

# SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: Takanori OKURA  
Kakuji TORIGOE  
Masahi KURIMOTO
- (ii) TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE OF  
INDUCING THE PRODUCTION OF INTERFERON- $\gamma$
- (iii) NUMBER OF SEQUENCES: 35
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: BROWDY AND NEIMARK
  - (B) STREET: 419 Seventh Street, N.W., Suite 300
  - (C) CITY: Washington
  - (D) STATE: D.C.
  - (E) COUNTRY: USA
  - (F) ZIP: 20004
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: Patent In Release #1.0, Version #1.30
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: JP 185,305/96
  - (B) FILING DATE: 27-JUN-1996
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: BROWDY, Roger L.
  - (B) REGISTRATION NUMBER: 25,618
  - (C) REFERENCE/DOCKET NUMBER: OKURA=1
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 202-628-5197
  - (B) TELEFAX: 202-737-3528

## (2) INFORMATION FOR SEQ ID NO: 1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 157 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

Tyr	Phe	Gly	Lys	Leu	Glu	Ser	Lys	Leu	Ser	Val	Ile	Arg	Asn	Leu	Asn	1	5	10	15
Asp	Gln	Val	Leu	Phe	Ile	Asp	Gln	Gly	Asn	Arg	Pro	Leu	Phe	Glu	Asp	20	25	30	
Met	Thr	Asp	Ser	Asp	Cys	Arg	Asp	Asn	Ala	Pro	Arg	Thr	Ile	Phe	Ile	35	40	45	
Ile	Ser	Met	Tyr	Lys	Asp	Ser	Gln	Pro	Arg	Gly	Met	Ala	Val	Thr	Ile	50	55	60	
Ser	Val	Lys	Cys	Glu	Lys	Ile	Ser	Xaa	Leu	Ser	Cys	Glu	Asn	Lys	Ile	65	70	75	80
Ile	Ser	Phe	Lys	Glu	Met	Asn	Pro	Pro	Asp	Asn	Ile	Lys	Asp	Thr	Lys	85	90	95	
Ser	Asp	Ile	Ile	Phe	Phe	Gln	Arg	Ser	Val	Pro	Gly	His	Asp	Asn	Lys	100	105	110	
Met	Gln	Phe	Glu	Ser	Ser	Ser	Tyr	Glu	Gly	Tyr	Phe	Leu	Ala	Cys	Glu	115	120	125	

Lys Glu Arg Asp Leu Phe Lys Leu Ile Leu Lys Lys Glu Asp Glu Leu  
 130 135 140  
 Gly Asp Arg Ser Ile Met Phe Thr Val Gln Asn Glu Asp  
 145 150 155

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1120 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: No

(iv) ANTI-SENSE: No

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
- (F) TISSUE TYPE: liver

(ix) FEATURE:

- (A) NAME/KEY: 5'UTR
- (B) LOCATION: 1..177
- (C) IDENTIFICATION METHODS: E
- (A) NAME/KEY: leader peptide
- (B) LOCATION: 178..285
- (C) IDENTIFICATION METHODS: S
- (A) NAME/KEY: mat peptide
- (B) LOCATION: 286..756
- (C) IDENTIFICATION METHODS: S
- (A) NAME/KEY: 3'UTR
- (B) LOCATION: 757..1120
- (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GCCTGGACAG TCAGCAAGSA ATTGTCTCCC AGTGCATTTT GCCCTCCTGG CTGCCAACTC	60
TGGCTGCTAA AGCGGCTGCC ACCTGCTGCA GTCTACACAG CTTGCGGAAG AGGAAAGGAA	120
CCTCAGACCT TCCAGATCGC TTCCTCTCGC AACAACTAT TTGTGCGAGG AATAAAG	177
ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA ATG	225
Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala Met	
-35 -30 -25	
AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA GCT GAA GAT GAT GAA AAC	273
Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala Glu Asp Asp Glu Asn	
-20 -15 -10 -5	
CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA TCT AAA TTA TCA GTC ATA	321
Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser Val Ile	
1 5 10	
AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT GAC CAA GGA AAT CGG CCT	369
Arg Asn Leu Asn Asp Gln Val Leu Phe Ile Asp Gln Gly Asn Arg Pro	
15 20 25	
CTA TTT GAA GAT ATG ACT GAT TCT TCT GAC TGT AGA GAT AAT GCA CCC CGG	417
Leu Phe Glu Asp Met Thr Asp Ser Asp Cys Arg Asp Asn Ala Pro Arg	
30 35 40	
ACC ATA TTT ATT ATA AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG	465
Thr Ile Phe Ile Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met	
45 50 55 60	
GCT GTA ACT ATC TCT GTG AAG TGT GAG AAA ATT TCA AYT CTC TCC TGT	513
Ala Val Thr Ile Ser Val Lys Cys Glu Lys Ile Ser Xaa Leu Ser Cys	
65 70 75	
GAG AAC AAA ATT ATT TCC TTT AAG GAA ATG AAT CCT CCT GAT AAC ATC	561
Glu Asn Lys Ile Ile Ser Phe Lys Glu Met Asn Pro Pro Asp Asn Ile	
80 85 90	
AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG AGA AGT GTC CCA GGA	609

Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Gln Arg Ser Val Pro Gly  
 95 100 105

CAT GAT AAT AAG ATG CAA TTT GAA TCT TCA TCA TAC GAA GGA TAC TTT 657  
 His Asp Asn Lys Met Gln Phe Glu Ser Ser Ser Tyr Glu Gly Tyr Phe  
 110 115 120  
 CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA CTC ATT TTG AAA AAA 705  
 Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys Leu Ile Leu Lys Lys  
 125 130 135 140  
 GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC ACT GTT CAA AAC GAA 753  
 Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe Thr Val Gln Asn Glu  
 145 150 155  
 GAC TAGCTATTAA AATTTTCATGC CGGGCGCAGT GGTCTACGCC TGTAATCCCA 806  
 Asp  
 GCCCTTTTGGG AGGCTGAGGC GGGCAGATCA CCAGAAGTCA GGTGTTCAAG ACCAGCCTGA 866  
 CCAACATGGT GAAACCTCAT CTCTACTAAA AATACTAAAA ATTAGCTGAG TGTAGTGACG 926  
 CATGCCCTCA ATCCAGCTA CTCAAGAGGC TGAGGCAGGA GAATCACTTG CACTCCGGAG 986  
 GTAGAGGTTG TGSTGAGGCG AGATTGCACC ATTGCGCTCT AGCCTGGGCA ACAACAGCAA 1046  
 AACTCCATCT CAAAAAATAA AATAAATAAA TAAACAAATA AAAAATTCAT AATGTGAAAA 1106  
 AAAAAAAAAA AAAA 1120

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 135 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
- (F) TISSUE TYPE: placenta

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 1..135
- (C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA TCT AAA TTA TCA 47  
 Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser  
 -5 1 5 10  
 GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT GAC CAA GGA AAT 95  
 Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe Ile Asp Gln Gly Asn  
 15 20 25  
 CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT AGA G 135  
 Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp Cys Arg Asp  
 30 35 40

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 134 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
- (F) TISSUE TYPE: placenta

(ix) FEATURE:  
 (A) NAME/KEY: exon  
 (B) LOCATION: 1..134  
 (C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

AT AAT GCA CCC CGG ACC ATA TTT ATT ATA AGT ATG TAT AAA GAT AGC	47
Asp Asn Ala Pro Arg Thr Ile Phe Ile Ile Ser Met Tyr Lys Asp Ser	
40 45 50 55	
CAG CCT AGA GGT ATG GCT GTA ACT ATC TCT GTG AAG TGT GAG AAA ATT	95
Gln Pro Arg Gly Met Ala Val Thr Ile Ser Val Lys Cys Glu Lys Ile	
60 65 70	
TCA ACT CTC TCC TGT GAG AAC AAA ATT ATT TCC TTT AAG	134
Ser Thr Leu Ser Cys Glu Asn Lys Ile Ile Ser Phe Lys	
80 85	

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 87 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: human  
 (F) TISSUE TYPE: placenta

(ix) FEATURE:  
 (A) NAME/KEY: exon  
 (B) LOCATION: 1..87  
 (C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GAATAAAG ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG	50
Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val	
-35 -30 -25	
GCA ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G	87
Ala Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala	
-20 -15 -10	

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 12 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: human  
 (F) TISSUE TYPE: placenta

(ix) FEATURE:  
 (A) NAME/KEY: exon  
 (B) LOCATION: 1..87  
 (C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CT GAA GAT GAT G  
Ala Glu Asp Asp Glu  
-10

12

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2167 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
- (F) TISSUE TYPE: placenta

(ix) FEATURE:

- (A) NAME/KEY: exon + 3'UTR
- (B) LOCATION: 1..2167
- (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GAA ATG AAT CCT CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA	48
Glu Met Asn Pro Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile	
85 90 95 100	
TTC TTT CAG AGA AGT GTC CCA GBA CAT GAT AAT AAG ATG CAA TTT GAA	96
Phe Phe Gln Arg Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu	
105 110 115	
TCT TCA TCA TAC GAA GBA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC	144
Ser Ser Ser Tyr Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp	
120 125 130	
CTT TTT AAA CTC ATT TTG AAA AAA GAG GAT GAA TTG GGG GAT AGA TCT	192
Leu Phe Lys Leu Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser	
135 140 145	
ATA ATG TTC ACT GTT CAA AAC GAA GAC TAGCTAT TAAATTTTCA TGCCGGGCGC	246
Ile Met Phe Thr Val Gln Asn Glu Asp	
150 155	
AGTGGCTCAC GCCTGTAATC CCAGCCCTTT GGGAGGCTGA GGCGGGCAGA TCACCAGAGG	306
TCAGGTGTTT AAGACCAGCC TGACCAACAT GGTGAAACCT CATCTCTACT AAAAATACAA	366
AAAATTAGCT GAGTGTAGTG ACCCATGCCC TCAATCCAG CTAATCAAGA GGCTGAGGCA	426
GGAGAATCAC TTGCACTCCG GAGGTGAGG TTGTGTGAG CCGAGATTGC ACCATTGCGC	486
TCTAGCCTG GCAACAACAG CAAACTCCA TCTCAAAAAA TAAAAATAAT AAATAAACAA	546
ATAAAAAATT CATAATGTGA ACTGTCTGAA TTTTATGTT TAGAAAGATT ATGAGATTAT	606
TAGTCTATAA TTGTAATGGT GAAATAAAAT AAATACCACT CTTGAAAAAC ATCATTAAAG	666
AATGAATGAA CTTTCACAAA AACAACAAA CAGACTTTCC CTTATTTAAG TGAATAAAAT	726
AAAATAAAAT AAAATAATGT TAAAAAAAT CATAGTTTGA AAACATTCTA CATTGTTAAT	786
TGGCATATTA ATTATACTTA ATATAATTAT TTTTAAATCT TTTGGGTTAT TAGTCCTAAT	846
GACAAAAGAT ATTGATATTT GAACTTTCTA ATTTTAAAGA ATATCGTTAA ACCATCAATA	906
TTTTTATAAG GAGGCCACTT CACTTGACAA ATTTCTGAT TTCTTCCAAA GTCAGTATAT	966
TTTTTAAATT CAGTTTGATC CTGAATCCAG CAATATATAA AAGGGATTAT ATACTCTGGC	1026
CAACTGACAT TCATCCTAGG AATGCAAGA TGGTTTAATA TCCTAAAATC AATTAACATA	1086
ACATACTATA TTAATAAAGT ATCAAAACAG TATTCTCATC TTTTTTCTT TTTTCACAAT	1146
TCCTTGGTTA CACTATCATC TCAATAGATG CAGAAAAAGC ATTTGACAAA ATCCAATTCA	1206
TAATAAAAT TCTCAAACTT GAAGGAGAAC ATCATAAAGG CATCTATGAA AAACCTACAG	1266
CTAATATCAT ACTTAACGAT GAAAAACTCA ATTATTTTAC CCTAAGATCA AGAATAATGC	1326
AAGCATGTCA GCTCTTGCAA CTCTATTCTA ACATTGTACT GGAGGTTCTA GCCAGAGCAA	1386
CCATACAATA AATAAAAAATA AAAGGCACCC AGATTAGAAA GGAAGTCTTT ATTTGCAGAC	1446
AACATGGTTT TTTATGCAGA AAACCGTCAG GAATACACAC ACATGTTAGA ACTAATAAGT	1506
TCAGCAAGGT TGCAGGTTGC AATATCAATA TGCAAAAAATA CATTGAAGGC TGGGCTCAGT	1566
GGAGATGGCA TGTACCTTTC GTCCCAGCTA CTTGGGAGG TGAGGTAGGA GGATCACTTG	1626
AGGTGAGGAG TTTGAGGCTA TAGTGCAATG TGATCTTGCC TGTGAATAGC CACTGCACTC	1686
GAGCCTAGGC AACAAAGTGA GACCCCTCT CCAAAAAAAA AAATGGTATA TTGGTATTTT	1746
TGTATATGAA CAATGAATGA TCTGAAAAACA AGAAAAATTC ATTACAGATG GTATTAAAAA	1806
AATAAAATAC AAATAAATTT AGCAAAATAA TTATAAAACT TGTACATCGA AAATTTCAAA	1866

GCACTCTGAG GGAAATTAAA GATGATCTAA ATAATTGGAG AGACACTCTA TGATCACTGA 1925  
 TTG3AAAATT CATTCAATAT TGTTAAGATA ACAATTGTCC CCAAATTGAT GCATGCATTG 1985  
 AATTTAGTCT TCATCAAAAT TCCAGCAGG3 TTTTTCGAGA AATTGACAAG CTGTACCCAA 2045  
 AATGTATATG GAAATGAAAA GACCCAGAA3 AGCAAATAAT TTTTAAAAA CAAASTTGA 2105  
 AAACCTTTTAC TTCTAATTT TAAACTTAC TATAAACCTA AAGTTATCAA GACCATTTAG 2165  
 T 2167

(2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1334 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: human
  - (F) TISSUE TYPE: placenta
- (ix) FEATURE:
  - (A) NAME/KEY: intron
  - (B) LOCATION: 1..1334
  - (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GTATTTTTTT	TAATTCGCAA	ACATAGAAAT	GACTAGCTAC	TTCTTCCCAT	TCTGTTTTAC	60
TGCTTACATT	GTTCCGTGCT	A3TCCCAATC	CTCAGATGAA	AAGTCACAGG	AGTGACAATA	120
ATTTCACTTA	CAGGAAACTT	TATAAGGCAT	CCACGTTTTT	TAGTTGGGGT	AAAAAATTGG	180
ATACAATAAG	ACATTGCTAG	G3GTTCATGCC	TCTCTGAGCC	TGCCTTTGAA	TCACCAATCC	240
CTTTATTGTG	ATTGCATTAA	CTGTTTAAAA	CCTCTATAGT	TGGATGCTTA	ATCCCTGCTT	300
GTTACAGCTG	AAAATGCTGA	TAGTTTACCA	GGTGTGGTGG	CATCTATCTG	TAATCCTAGC	360
TACTTGG3AG	GCTCAG3CAG	GAGGATTGCT	TGAGGCCAGG	ACTTTGAGGC	TGTAGTACAC	420
TGTGATCGTA	CCTGTGAATA	G3CACTGCAC	TCCAGCCTGG	GTGATATACA	GACCTTGTCT	480
CTAAAATTAA	AAAAAAAAAA	AAAAAAAAAC	TTAG3AAAGG	AAATTGATCA	AGTCTACTGT	540
GCCTTCCAAA	ACATGAATTC	CAAAATATCAA	AGTTAG3CTG	AGTTGAAGCA	GTGAATGTGC	600
ATTCTTTAAA	AATACTGAAT	ACTTACCTTA	ACATATATTT	TAAATATTTT	ATTTAGCATT	660
TAAAAGTTAA	AAACAATCTT	TTAGAATTCA	TATCTTTAAA	ATACTCAAAA	AAGTTGCAGC	720
GTGTGTGTTG	TAATACACAT	TAAACT3TGG	G3TTGTTTGT	TTGTTTGAGA	TGCAGTTTCA	780
CTCTGTCAAC	CAGGCTGAAG	T3CAGT3CA3	T3CAGTGGTG	TGATCTCG3C	TCACTACAAC	840
CTCCACCTCC	CACGTTCAAG	CGATTCTCAT	G3CTCAGTCT	CCCGAGTAG3	TGGGATTACA	900
GGCATGCACC	ACTTACACCC	G3CTAATTTT	T3TATTTTTA	GTAGAGCTGG	GGTTTCACCA	960
TGTTGG3CA3	GCTG3TCTCA	AA3CCCTAAC	CTCAAGTGAT	CTGCCTGCCT	CAGCCTCCCA	1020
AACAAACAAA	CAACCCACAA	GTTTAATATG	TGTTACAACA	CACATGCTGC	AACTTTTATG	1080
AGTATTTTAA	TGATATA3AT	TATAAAAG3T	TGTTTTTAAC	TTTTAAATGC	TGGGATTACA	1140
GGCATGA3CC	ACTGTGCCAG	G3CTGAACTG	TGTTTTTAAA	AATGTCTGAC	CAGCTGTACA	1200
TA3TCTCCTG	CAGACTGGCC	AA3TCTCAAA	GTGGGAACAG	GTGTATTAA3	GACTATCCTT	1260
TG3TTAAATT	TCC3CAAAATG	TTCTGT3CA	AGAATTCCTC	TAACTAGAGT	TCTCATTTAT	1320
TATATTTATT	TCAG					1334

(2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 4773 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: human
  - (F) TISSUE TYPE: placenta
- (ix) FEATURE:

(A) NAME/KEY: intron  
(B) LOCATION: 1..4773  
(C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GTAAGACTGA	GCCTTACTTT	GTTTTCAATC	ATSTTAATAT	AATCAATATA	ATTAGARATA	60
TAACATTATT	TCTAATGTTA	ATATAAGTAA	TGTAANTTAA	AAACTCAAAT	ATCCTCAGAC	120
CAACCTTTT	TCTAGAACAG	AAATAACAAG	AAACAGAGAA	CCATTAAAGT	GAATACTTAC	180
TAAAAATTAT	CAAACCTTTT	ACCTATTGTG	ATAATGATGG	TTTTTCTGAG	CCTGTACAG	240
GGGAAGAGGA	GATACAACAC	TTGTTTTATG	ACCTGATCT	CCTBAACAAT	CACTCTTTAT	300
ATAAATAATA	ATGTAGAATA	CATATGTGAG	TTATACATTT	AAGAATAATA	TGTACTTTT	360
CAGAAATGAT	TCTGCTATGA	AGAAATGAAG	TAATTATCCT	TCTATATTTT	TACACCTTTG	420
TAAATTATGA	TAATATTTTA	ATCCCTAGTT	GTTTTGTGTC	TGATCCTTAA	CCTAAATCTT	480
AGACACAAGT	TTCAGCTTTC	ATTTATGTGA	TCTTATTTTT	AATGTTAATC	TAATTGAATA	540
AAAGTTATGA	GATCAGCTGT	AAAAATTAATG	CTATAATTAT	CTTCAAGGCA	GATATAAAGT	600
ATTTCTGTC	TCTACTTTTT	CTCTATTATT	CTCCATTATT	ATTCTCTATT	ATTTTTCTCT	660
ATTTCTCTCA	TTATTGTTAG	ATAAACCACA	ATTAACTATA	GCTACAGACT	GAGCCAGTAA	720
GAGTAGCCAG	GGATGCTTAC	AAATTGACAA	TGCTTCAGAG	GAGAATTCTA	TGTCATCAAG	780
ACTCTTTTTT	ATGTGAATAT	TGCCAATAAA	TACTGCTTTT	CATGCCCAAC	CACTGCCCAC	840
TGAAGAGACG	TTAGGATATG	ACCTTATATG	ATGTATCCAA	GGGCAACTTG	GTAGGAGAAA	900
AAAAGCCACT	CTAAATATATA	ATCCAAATTA	GAACAGTGCA	TATGCAACAG	ATACAGCCCC	960
CAGACAATTC	CCTCAGCTAT	CTCCCTCCAA	CCAGAGTGCC	ATCCCTTCAG	GTGACAAATTT	1020
GGAGTCCCAA	TTCTAGACCT	GACAGGCAGC	TTATTTATCA	AAATAGCATA	AGAGCCCTGG	1080
GATGGAAGGG	TAGGCTGAAA	AGGCTTAAGC	ATGCTGTTAC	TGAACAACAT	AATTAGAAGG	1140
GAAGGAGATG	GCCAACTCA	AGCTATGTG	GATAGAGGAA	AACTCAGCTG	CAGAGGCAGA	1200
TTCAGAAACT	GGGATAAGTC	CGAAGCTACA	GCTGGATTCT	TGTTGAGGGA	GACTGCTGAA	1260
ATGTTTAAAG	AGATGGAAT	AATGCTTGG	ACTTAGTAGG	AACTGGGCAA	ATCCATATTT	1320
GGGGAAGCCT	GAGTTTTATT	CAATTTTCA	GGCCTTTTTA	ATAAATAAGA	ATGTGGCTTG	1380
GGGTGCTTGG	TCACACCTGT	AATCCCAACA	CTTTGGGAGG	CGAGGGGGGG	CGGATCCACT	1440
GAAGTCAGGA	GTTCAAGACC	AGCTTACCA	ACATGGAGAA	ACCCCATCTC	TACTAATAAT	1500
ATAAATTTAG	CTGGGCGTGG	TGGCATATGC	CTGTAATCCC	AGCTACTCGG	GAGGCTGAGG	1560
CAGGAGATTC	TTTTGAACCC	GGAAGGAGAA	GCTTGGGATG	AGCTTAGATC	GTGGCATTTG	1620
ACTCCAGCCT	GGGCAACAG	AGCAAAATC	GCTCTCAAAA	AAAAAATAAA	AAAGTGAATA	1680
TTAACCAGAG	GCATTAGCTT	AATAATTTAA	TACTGTTTTT	AGGTAGGGGG	GGGGTGGGCT	1740
GGAGAGATTC	TGTGTAAATG	AGGGAATGTG	ACATTTTAA	TTGATCAGCA	TCATAGCAAA	1800
TCTGCTTCTG	GAAGGAACCT	AATAAATATT	AGTTGGAGGG	GGGAGAGAGG	TGAGGGGTGG	1860
ACTAGGAGCA	GTTTTAGCCC	TTGCTTTTAA	TGGCTTTTTG	TGGCAATTAAT	AAGGATCTTA	1920
GCAATGCTTA	TAAAGTGGG	CTAGGTTCTA	GATAAATAAA	TACAACAGCC	CAGGCACAGT	1980
GGCTCATCTC	TATAATCCCA	GCCTTTGGGG	AGGGCAAGGG	GAGTGTCTTA	CTTGAGATCA	2040
GGAATTTAAG	ACCAGCCTGG	CCAGATGGG	GATACTCTGT	CTCTACTAAA	AAAAATACAA	2100
AAATTATCTA	GGCATGGTGG	CATGCACTTG	TAATCCCAAT	TACTGCTGAG	CCTGAGGCGAG	2160
AGAAATGCTT	TGAAGCAAGG	AGGTGTAGGG	TGATGTAGAG	TGATATGGA	CCACTGCACT	2220
CCAGCTCTGG	CGACAGAAAT	AGACTTTGTC	TCAAAAAAAG	AAAAAGATAC	AACAGGATAC	2280
CCTTATCTGT	TCACCTTTCA	CTGTTGATTA	CTAGCTATAA	AGTCTATATA	AGTCTTTTGG	2340
TCAAGAGACT	TGACAAACAT	AAGAGGGATT	TGCTTTTGA	GCTTACTGTC	AGATCTCTGT	2400
TCATATATAT	ACATATACAT	GTATATATGT	ATCTATATCC	AGGCTTGGGG	AGGCTTCCCT	2460
CAGACTTTCT	AGTGCACTTG	GGAGATGTTA	GCTCAATATC	AACTTTCCCT	GGATTCAAGT	2520
TCAACCCCTT	CTGATGTAAA	AAAAAATAAA	AAAAAGAAAG	AAATCCCTTT	CCCTTTGGAG	2580
CAGTCAATTT	TCAAGAGGTG	GGGCTTTTCA	AGTTGGGGGT	TCTCCAAAGT	CATTGGGATT	2640
GCTTTCAAT	CCATTGCTTA	TGTACCTTCC	CTATGATGTC	TGGGAGTGGT	CAACATCAAA	2700
ACTAGGAAAG	CTAGTGGCCA	AGGATGTCTT	TAGCTCATAT	CTGAAATTTG	CAATAAGTGT	2760
GATTAAATAG	ATTGCTGTGT	CTACCTATCC	ACATCTCTGG	TTTCAACTGT	AACTTTCTTT	2820
TTTTCTTTT	TCTTTTTTTT	CTTTTTTTTT	GAAAGGAGGT	CTGCTCTGCT	CGGCAAGGCT	2880
AGAGTGAAGT	GGCAGGATCT	CAGCTCACTG	CAGGCTGTC	CTCCCGGCTT	CAGGCTATTC	2940
TGCTGCTTCA	CCCTGCCAAG	CAGCTGCGAC	TACAGGCTTC	TGCGAGCTAG	CCCACTTAAT	3000
TTTTGTATTT	TTTAGTAGAG	ACGGGGTTTT	ACGGGTGTTG	CGAGGATGGT	CTGATCTCTC	3060
TGAATTTTGT	ATCCGCGCGC	CTCAGCTCTC	CAAAATGATG	GGATTACAGG	CGTGAAGCAT	3120
CGCAGCCGGC	TCAACTGTAA	CTTTCTATAT	TGCTTCACTT	TCCCTGTAA	TGTTACTAGA	3180
GCTTTTGAAG	TTTTGGCTAT	GGATTATTTT	TGATTTATAT	ATTAGATTTC	AGATTATGTC	3240
CAAAATGATG	CCACAGCTTT	AGGCTCTCTT	CTAAATCTGT	ATATTGTAGA	CAGCTGAGAA	3300
AGTGCTGTGC	AATAGGGGAA	CTAGTTTATA	CTTTCACTAA	CTTAGGAGCC	AGACTTGTGT	3360
ATAAGGAACA	AAGGTCAAGA	GTTATGACTA	CTGATCTCAC	AACTGATTGA	GAATTTGGAG	3420
ATAACCCGCT	GACCTCTGCC	ATCCAGAGTC	TTTCAGGAT	CTTTGAAGGA	TGAAGAAATG	3480
CTATTTTAA	TTTGGAGGTT	TCTCTATCAG	TGCTTATGAT	CATGGGAATC	TGTGCTGCCA	3540
TGAGGCCAAA	ATAAGTCCCA	AAACATCTAC	TGCTTCCAGG	ATTAACATGG	AAAGAACTTA	3600
GCTGCTGCCC	ACATGTTCTG	ATCCATCTCT	CAAAATAGAC	ATGCTGCACT	AACAAGAAAA	3660

GTGCAGGCAG	CACTACCACT	TGGATAACCT	GCAAGATTAT	AGTTTCAAGT	AATCTAACA	3720
TTTCTCACAA	GGCCCTATTG	TGTGACTGAA	ACATACAAGA	ATCTGCAATT	GGCCTTCTAA	3780
GSCAGGGCCC	AGCCAAAGAG	ACCATATTCA	GGACAGAAAT	TCAAGACTAC	TATGGAAGTG	3840
GAGTGTCTTG	CAGGGAAGAC	AGAATCAAGG	ACTGCCAACT	GAGGCAATAC	AGTAGGCTTA	3900
CACAGGAAGC	CAGGCTTAG	CCCTACAACA	ATTATTGGGT	CTATTCACTG	TAAATTTTAA	3960
TTTCAGGCTC	CACTGAAAGA	GTAAGCTAAG	ATTCTCTGGC	CTTTCTGTCT	CTCTCAGAGT	4020
TGGCTCAGAA	ATGAGAACTG	GTCAGGCCAG	GCATGGTGGC	TTACACCTGG	AATCCCAAGC	4080
CTTTGGGAGG	CCGAAGTGAG	AAGGTCACTT	GAGGCCAGGA	GTTCAGGACC	AGCTTAAGCA	4140
ATAAAGTGAG	ATACCCCTCT	ACCCCTTCTC	TACAAAAATA	AATTTTAAAA	ATTAGCCAAA	4200
TGTGCTGTGT	TATACTTACA	GTCCCACTA	CTCAGGAGGC	TGAGGCAGGG	GGATTGCTTG	4260
AGCCCAAGAA	TTCAAGGCTG	CACTGAGCTA	TGATTTCAAC	ACTGCACTTC	TGGCTGGGCA	4320
ACAGAGGAG	ACCCTGTCTC	AAAGCAAAAA	GAAAAAGAAA	CTAGAACTAG	CCTAAATTTG	4380
TGGGAGGAGG	TCATCATCGT	CTTTAGCCGT	GAATGGTTAT	TATAGAGGAC	AGAAATTGAC	4440
ATTAGCCCAA	AAAGCTTGTG	GTCTTTGTCT	GAAGCTTACT	TAATCTTGAG	CAAAATGTGA	4500
CACCACTCAA	TGGGAGAGGA	GAAGAAGTAA	CTGTTTGATG	TATAGGGAAG	AACTAGAGGC	4560
CTGGAAGTGA	ATATGCATCC	CATGACAGGG	AGAATAGGAG	ATTCTGAGTT	AAGAAGGAGA	4620
GGAGGTCACT	ACTGCTGTTC	AGAGATTTT	TTTATGTAAC	TCTTGAGAAG	CAAACTACT	4680
TTTGTCTCTG	TTGTAATAT	ACTTCAAAAC	AACTTCATA	TATTCAAATT	GTTTATGTCC	4740
TGAATAAATT	AGGTAATGTT	TTTTTCTCTA	TAG			4773

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8835 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
- (F) TISSUE TYPE: placenta

(ix) FEATURE:

- (A) NAME/KEY: intron
- (B) LOCATION: 1..8835
- (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

GTAAGAAATA	TCATTCTCT	TTATTTGGAA	AGTCAGCCAT	GGCAATTAGA	GGTAAATAAG	60
CTAGAAAGCA	ATTGAGAGGA	ATATAAACCA	TCTAGCATCA	CTACGATGAG	CAGTCAGTAT	120
CAACATAAGA	AATATAGGCA	AAGTCAGAGT	AGAATTTTTT	TCTTTTATCA	GATATGGGAG	180
AGTATCACTT	TAGAGGAGAG	GTTCTCAAAC	TTTTTGCTCT	CATGTTCCCT	TTACACTAAG	240
CACATCAGAT	GTTAGCATAA	GTAACATTTT	TAATTAAGAA	TAACTATGTA	CTTTTTTAA	300
AAACAAAAAA	AGCATAAGGA	GTGACACTTT	TTTATTTTTT	CAAGTGTTTT	AACTGGTTTA	360
ATAGAAGGCA	TATAGATCTG	CTGGATTCTC	ATCTGCTTTG	CATTGAGACT	ACTGCAATAT	420
TGCACAGAA	GCAGCTCTCT	GTAACCTCTG	TTGTACACTC	ATGAGAGAA	GGGTGAAAA	480
GACAAATTAC	GCTTTAGAA	TATTAGAAAT	AGCTTTCACT	TTAGGAACTC	CTGAGAATT	540
GCTGCTTTAG	AGTGGTAAGA	TAAATAAGCT	TCTCTTTAAA	CGGAATCTCA	AGACAGAACT	600
AGTTACATTA	AAAGCAAAAC	AAAAATTTGC	CCATGGTTAG	TCATCTTGTT	AAATCTGCCA	660
CACTCTTTGA	CTGGGCTACA	ATTGGATAAT	ATAGCATTC	CCGAGATAAT	TTTCTCTCAC	720
AATTAAGGAA	AGGGCTGAAT	AAATATCTCT	GTTTGAAGTT	GAATAACAAA	AATTAGGACC	780
CCCTAAATTT	TAGGCTCCCT	GAAATTCGTC	TTTTTGCCCTA	TATTCAGCTA	CTTTACGTTT	840
TATTAATCTT	TCTTTTCAGG	CAGGTGCACT	AGCTCATGCC	TAGAACTCTA	GTCAGGCTCT	900
AGCCCAAGAA	TTTGAAGCCA	GCCAGGGCAA	CACAGTCTCT	ACAAAAAAT	AAAAAATTAC	960
CTGGCTGTGT	TGGTGCATGC	CTGTAAGAA	ACTCAGGATG	CTGAGGACTG	CTTGAGGCCC	1020
GGATAGCCAA	ATCTGTGGTG	AGTTCAAGCA	CTAAACAGAG	CSAGACTTTT	TCAAAAAAAC	1080
AAACAAAAAA	ACAAACAAAC	TTCTTTCAAA	ATAACTTTTT	ATCTGCAATG	TTTTCTTATT	1140
GCTGTGAGTA	TTAAATTTAC	TCTTTTCACT	GATTTCCAAA	GCCCTCCATA	ATCTAATCCG	1200
ACTTTACCTT	GTGTTCACTG	CAAAATAGCA	GGACTGTTCC	ACTACAAATC	AAAAATCACA	1260
GCTTGCGGTC	AGTGGCTCAC	TCCTGTAATC	CCAACACTTT	GGAAGGCCAA	GSCAGGCTGA	1320
TTGCTTCAGC	TCAGGAGTTC	AAGACCAGCC	TGGGCAACAT	GGCAAAAAAC	CTGTCTCTCC	1380
AAAACATACA	AAAATTAGCC	AGATGTGGTA	GATGTGCCCT	GTAGTCCCAA	CTACTCAAAA	1440
GGCTAAGGCA	AGAGGATCAC	TTGAGCCGAG	GAGGTCAAAG	CTACAGTGA	CCATGTTTAC	1500
TGTGTCACTG	CACTCCAGCC	TGGGTGATAG	AGCAAGACCA	TGTCTCAAAA	AAAAAATAAA	1560

GAAGAGAAAA	GAAGAAAAACA	TGGTCTATT	CAGTTCCACC	CCAGCACAAC	ATTGTTTTBA	1620
TTATCACATA	AATGCTGGTC	CATTGCOOTC	TCTATCTATT	CAATCTTTA	AGCATTCTTT	1680
GAGATTCAA	TCAATTCTCC	TTTTCAACT	AGGCCATTTA	AACTACATCA	GTTCACATTT	1740
GATTTTCTT	CTTTGAGTCT	ATAGACTCAA	AAACAAAAAC	TTAAAAACTT	ATTTTTTAA	1800
TTTTCTGCTA	CTCTCACTTC	TTCAACAATC	ACATACACGC	ATTDATAATA	ABATGGCAGA	1860
ATGTTCAA	ATAAAATGAT	TTATAGAACT	GAAAASTTAS	GTTTTGATCT	TBTGCTCTC	1920
AAGATGACTA	CCTACCTGAT	CTCAGSTAA	TAATTATGTA	GCATGCTCCC	TCATTTTCATC	1980
CCATACTAT	TCAACAGGAT	TGAAATTCOA	CAGCAAGGAT	AAACATAATC	ATAGTTGCTT	2040
TTCAAATTCOA	AGGCATTTTA	ACTTTTAATC	TAGTAGTAGT	TTTGTGTTG	TTGTTGTTGT	2100
TTGAATGGA	GCCCTGCTGT	GTCACTCAAG	CTGGAGTGCA	GTGCAACBAA	CTCGCTCTAC	2160
TGCAACCTCT	GCCCTCATG	TTCAATCAAT	TATTCTGCT	CAGTGTCCCA	AGTAAGCTG	2220
ACTACAAG	ACATGCCACC	ATGCTGCTG	AATTTTTGTA	TTTTTAAGTA	AAACAGGCT	2280
TCACCATGTT	GGCCAGGCTG	GTCTGCAACT	CCTGACCTCA	AGTGATCCAG	CCGCTCTGCT	2340
CTCCCAAGT	GCTGGGATTA	CAGGCATAAG	CCACCGTGCT	CAGCCTAATA	GTATGTTTTT	2400
AAACTCTTAG	TGGCTTAACA	ATGCTGCTG	TATAATAAAT	ATGCCATAAA	TATTTACTGT	2460
CTTAGAATTA	TGAAGAAGTG	GTTACTAGGC	CGTTTTGCCAC	ATATCAATG	TTCTCTCTCT	2520
ACAGCTTTAA	TTAGAGTCTA	GAATTGCTAG	TTGGTAGAGC	TGGAACAGAC	CTTAAAGATT	2580
GACTAGCCAA	CTTCTCTGTC	CAAAATGAGG	AACTGAGAC	CTTAAATTA	AGTGACTTGC	2640
CCCAGACAA	ACTGSAACTC	ATGTGTCTA	ATTTCCATCA	TGAAATTTCTA	CCATTCTACTA	2700
GCTCTCTGCT	AGTTGTCAAA	GTATGCTATA	ACTAAATTTT	TATGCTCTGT	TTAAAGACAA	2760
AATTGTCACT	GCTTACTCCT	GGGAGGCTCT	TTCTGAGGTC	GTTTATAACT	CTTAAAAAAA	2820
AAAAAGTCAG	TAGTCTGAGA	ATTTTAGAGC	AAATAGTCAA	AGCATTTTTA	TCCAATGGAT	2880
CTATAATTTT	CATAGATTAG	AGTTAAATCA	AAAGAACACG	GATGAGAAAG	GAAGAGSAAA	2940
ATTGAGSAGA	GGAGSAATG	GGATGAGAAC	ACACTACTTG	TAATCAGTCA	TAGATGTACT	3000
GAGAACTAAC	AAGAGGAAT	GTAAAGAAAT	AGAAATGAG	AATTCAAAAT	CAAGACATGA	3060
AATAAAAGA	AACCTACTAG	GAAAAATGGA	GAAGACATTA	GAAAAATTAT	TCTATTTTTA	3120
AAATTTCTGT	TTCAAGCTTC	CCTCTCTGTC	TTCTCTCTTC	TCATTGGTTT	TCAGGTGAG	3180
GGAATTTTA	AGATGGGAAA	AATATATATA	TTCTACACAT	CCCTTTCTAC	GCTGTTGCTA	3240
TGGTAAACA	GTTTATCATA	GCAAACTTTT	ATTCATACAA	CATTTATTGA	GTTTCTACTG	3300
TGTGGTAA	TCTTTCCAG	TGTTGAAAT	TCAGGGGAAA	AAAGACAACT	CATTGTCTTA	3360
AAACTCAGAT	GAAGCTGAA	CAGACCTATT	TTTAATCATA	GTAATCTCAA	TTTAGGCTAG	3420
TAGAGCTAT	TTAAGAGGCA	TGAACAGGTC	TGAAGGAGGT	AGGACTCTGA	GGAGAGATA	3480
GTTAGCTAG	AATGAAGAG	CAGAGAGTT	TTCTTAGAGG	AACTATTAAA	GCTGGGASTT	3540
ACGGGATGAA	AGATGAGGCA	GCGTTTCTAG	GCAAAAAAAA	AAAAAGGCA	GGGAGGGG	3600
AAGTTCTGCT	CTGGTACAGA	GAATTAATGT	GCAACAAATG	GAGGAGAGTC	TGAGAGCAAG	3660
AAAGCCAGT	AGAAGATAT	TAAATAGAA	GATGCCAGGG	GTAATGAGGG	CTTGATTTAA	3720
AACAGTGCTG	TTGAGATGG	AGAGGAGATA	CCAAATTTCT	GAGACATTTG	TGAGTTAGAA	3780
CCTACAGTAT	TTATCAGACA	AAGGAAAGAT	TAGACAAAGG	AGTTAAGAA	GACTCCCGAG	3840
TTTCACTTTT	GGGCAAGTAA	CTAGGACATG	TTTTGMAAAG	TAAATGTTATG	GATCTCTTAC	3900
CATTGSAACT	ATGTATGTGG	AGCCTAATTA	AAATTTGTAC	ATGTATATAA	CTCTCCCGCC	3960
ACCAACASTA	ACTACTTCCC	TAACTCTCTA	CTTTGTAGCC	AGACTTCTTA	AAAGAATAGT	4020
TTTAGTCTAC	TGTCTTTACT	TTTCCCTCTC	CATTCTGTCT	TAGATATTGG	TCCACCTACC	4080
ATCTGCTGCT	TCCACTTTAC	CCAACTCTTT	CTAAGGTTGC	CCAAACTTTC	CTAATTTCTA	4140
AATTCAATGA	ACAAGTTTAA	GCTTATATGT	AAATTAGGAG	CTCTACAGTT	TGATTTGAG	4200
CAGCCCTCTC	TGAAGCCCTT	TCTTTCTGCA	CTCTCTGAG	ACATCTCAGA	TTTACAAAAC	4260
TGAAGTAAT	ATTTTACACT	TGAGCTGTAT	TTTCTGTTCT	CTTTCTTCTA	GAATGAGGTA	4320
ACCACTCAAC	AAATTGCCCA	AGCCAAAGAC	TACAAAGTCA	TCTTCAGTTC	CTCCTCTCTC	4380
TGTTTGAACC	ACAACAGATC	AGCTGAGAAA	TCCCGCTGTT	TAGTATCTCT	TGAATTCATT	4440
ACCTTAATTT	ATAGCTCTCAT	CAACTCTTAA	TCTGTTAAAT	TACTTCAGTA	GTTGTTCTCT	4500
GACTCTGTCT	CAATCTTGT	CAATGAGGTC	CATCTTTTTG	TCTTGTGTTG	TGGTGGTGT	4560
GTGACAGAG	TTTCTCTTTT	GCTTCTCAGG	CTGAAGTCTA	GTGAGGCACT	TCATTCGAAC	4620
CACAGCTCTC	TGGTTTAAAG	CAGTTCAACC	TCTTGAGTAG	CTGGGAGTAC	AGGTATCTGC	4680
CATACAGCTC	AGCTAATTTT	GCTTTTCTAG	TAGAGACAGG	GTTTCAACAT	GTTTCTCAGG	4740
CTGCTCTCAA	ACTCTCTGAC	TCAAGCAATC	CAGCTCACTC	AGCTCTCCAA	AGTCTCTGGA	4800
TTACAGGCTAT	GAGCTCTGCT	AGAGGAGTCA	GATCCATGTT	TTATGTGCT	TCTAGAGTGA	4860
GTTTITTAATA	CACAAATTTG	ACCATATCTT	TCTTCAATTT	AACTCAATAT	TTTTTTTTTT	4920
AGGAAAAAAC	AGTTCAAACT	CTTTAGTCTG	CTTACACAA	GCTTTTGTAG	TCTGACTCTT	4980
CTTTTCAAG	TTTCTATCAA	GTATCTGCTA	AGTTACATTT	TATGTGAAT	GAATTAAGCA	5040
AGCTATATAA	AATATAGTTT	TATATGCTCA	AAATGGAAAT	AATGTTAACT	CTTCTCAATA	5100
GTTTATCTAA	AATGACATAA	TTTCAAGGCT	GTGAGGTCAA	ATGATTTATA	AACTTTAAAC	5160
ACTATTGCTA	CATGCAAGTG	TCTTTTATAC	TTTGTAGAA	TATCTCTTTC	CATGTCATTA	5220
TTATCTAAAT	TAGACTTTAA	ATAACTCAGA	AGTTCTTCTG	ACATACAGGT	TATTATTTGT	5280
CTTTTAAAC	ATAATTTTAA	ATAATTTTAT	ATATGATAAT	GTTATCCAG	TGCTAAGGGA	5340
TGTATTGTTA	CTGCTGTGCA	AAAAAAAAC	AAAAAAAAC	TCCAAATAAA	TATGTTGAAA	5400
CCAGGTTTAT	ATGCAAGAAA	ACAATATTAA	AAAGGCTCAA	GTACCACCAT	AATAGGCTGT	5460
GTGAGAGCG	CAGGCTACAA	AACACTASTA	ATAATGCTGA	GAAAGTTGAA	AAAAAGAGA	5520
AAAGCAACAAT	ATGCTTTTGT	TGTTGTAGGT	TTATGTACTC	CAAGAAATATC	TCTCTCAAA	5580

CTTTTAC3TT	TTTTCCAAAG	AAAASTTAA	TTTGGCTGGG	CGCAGTGGCT	CTTGCTGTGA	5640
GTCCCCAGCT	TTGGGAGGCT	AAGGCGGCA	GATCACCTGA	GGTCAGGAGT	TTGAGACGAG	5700
CCTGACCAAA	AATGAGAGAA	CCCGGCGGCT	TCACTACTAA	AAGAATACAA	AATTAGGCCG	5760
GGCAGAGTGG	CTTACCCCTG	TGATCCAGG	ACTTTGGGAG	GGCGAAGCAG	GAAGATCACC	5820
TGAGGTCAAG	AGTTGAGAG	CAGGATGGA	GAAACCCGTC	TCTACTAAAA	ATACAAAAAT	5880
AGCGGGGCT	GGTGGTGCAT	GACTGTAATC	CCAGCTACTC	AGGAGGCTAA	GGCAGAGAA	5940
CACTTGAACC	CAGGCAAGT	AGGTTGCAAT	GAGGCGAGAT	CGTGCCATTG	CACTCCAGCC	6000
TGGGCAACAA	GAGGGAAGCT	CTGTATCCAA	AAAACAAAAG	AAAAGAAAAG	GTAACTTTGA	6060
ACTATGTGAG	ATCTTTAGAA	ATGCATTCTT	TCTGTAAAA	GTGACTACAT	TTGCCCTTAT	6120
TATGGTAAAA	ATGTTGAGG	CTCAACAAAC	CCATATTTTC	TGGGTCTCCC	CGTGGCTAG	6180
CCTTTGTTC	CATTGCTTCT	TCTTGGTGA	AGCTCTTCCT	CTGGGCTTGA	AAATGCTTGC	6240
TTCTCTTTCA	AGTAGGACAA	GTCATCACTT	TCTGTGGTAA	CCTTCTCCAG	CATCATCAAA	6300
CAAAAAGAA	GATCTCTTTC	TAAATTCAGC	TCTTACGTCA	TTCATTACAT	TATTTTGTAA	6360
CTCTTTATAG	ATTCTTCTCT	CCCACTAGAC	TCTGAGTCAC	TGGAGAGTAG	GAGCCAACTC	6420
TCATTCATCT	GTGGTTTGGT	CAGCTACTGG	CCACATTCCCT	GATGCATAGT	TAATGCTCAA	6480
ACCTTAAGCT	GTGAATCAGC	TCAATATTGG	TCCTTCTCTA	AATCCATTCA	CTCATTGACT	6540
AACTATGTAC	TCAAAATAGT	AAACACCACT	AATTTAATCC	AATTCCTGCC	CATACTGCTT	6600
GATACATTT	AGGTGAATTA	GTTTGATAAA	TATGTGTGTA	TTACATAATA	TTAAAGTATG	6660
TACAGAGAG	CATGCTAATC	ATAATTCACA	ACTGATAACT	AATCAAACAT	AAATGCTCTC	6720
AGGTTAACAA	ATGTCTGCC	TCTCAGTTAA	TGCAGTCATT	AACAAACACC	TTCTGATGCT	6780
GATAATAGGG	CCCTGTTTCAG	CAATGAAGCC	ATAAAGGTGA	ATAAAGAACA	TGCCCTCGTG	6840
GAGCTCAGAG	CCTAGTCACT	ATTGTTCTGA	TTTTTAATAT	TAATGTTGGT	TTGGGTTTTG	6900
GTGAAATATG	TTTAGACTTA	TCTTAGTGAT	CTTTTCATCC	TTTGCTATAT	TATTTTTCTC	6960
TAAGAGTCTT	CCCTATCCCC	TCTTTTAAAA	AAC TAGGTGA	TAATTTCTAA	TTGTAAATTT	7020
AAATATTATA	AATAGCTTAT	AAAATTTAAT	ATTTATAATA	TTTAAATGTT	TGATAAATAT	7080
TTAAATTTTA	TAATATTTAA	ATGTTTATTT	AAATTCATTT	GTACATCAGT	TTTTATTTTA	7140
TTTAAATGTC	TTGGCCAGGC	ATGTTGCTG	ACACCTATAA	TCCCAGAACT	TTGAGAGGCC	7200
AGGTCAGGCA	AACCATTTGA	GCTCAGGAGT	TTGAGACCA	CCTGGGCAAC	GTGGTGAAC	7260
CCTGTCTCTA	CCAAACATAT	GAAACCTTAT	CTGSETGTGG	TGGCACGCAT	CTGTGGTCCC	7320
AGATGGBAGT	CCCAGGCTAA	GATGGGAGAA	TGGCTTGAAC	CCAGGTGAGA	GGGGTGGGGT	7380
GGATGTTGCA	GTGAGCTGAG	ATGCTTCCAC	TGCATCCCAA	CCTGGGTGAC	AGAGTGAGAC	7440
TCCATCTCAA	AAAAAATAAA	TGTTATCTAA	ATAAGATATA	TTTAATAACT	GTTCCGACTT	7500
AGATGAGCAT	AGGGAATCAA	ACCTATATAA	AACATACAAA	TAAGGCCTGG	GTACAGTGAC	7560
TCATGCCCTG	AATCTCAGGC	ACTTTTGGAG	GGCAAAATTA	TACAAAGTTA	GTTGTATAAC	7620
ACCAACTAAC	AACATATTTG	GCGTTAAGCT	AATTCAGATT	AATTTTTTTT	AAACTGAGTT	7680
TTAAATTCCT	GCTTACTCTA	CCATACATGC	TAAGCCTCAT	ATTATGCTAG	AAAAATTTTG	7740
AGCACAGATT	TATGAATACT	CTCCTGCTAA	CCATTTAATT	TTTAAACAAA	TTTTAATGCA	7800
GTATATATCT	GCCTTTTTAC	CAACACATTA	AATAATAAGA	TCTACTGTGA	GGACTAAATT	7860
TCTGTAAAT	CMAAGTAGTA	ATGAGTTTAA	ACCATGTCTC	AAGATCTCTG	CAATAACTGT	7920
AGCACACAG	AAAAATAGTA	TTTCTATTAA	TGACAGAGTC	ACAAGTACTA	CTAATAATAC	7980
TGTGGTTTTG	TTCTCTCAAC	TATCTATGGG	AGGAATGCTA	AATTTTCAGAG	GTTGGTGAAA	8040
ATACATGTCT	ATTTTTTTTC	CCATCCCAAGT	TCACAGATTT	CTCACACTGA	GAACCTCTAT	8100
TCCATACACA	AATTTCTGAA	GCTCTCACAC	CGTATTGGAA	GAAGGGCAGA	AAGGAAAAGC	8160
AAATGGAGAG	ATTTAAATTT	TTTTCAAAATC	CTGTATCCCT	TGATTTTACA	GCAAGATTGT	8220
ATTTATGTAT	TACTTCTCTT	AAAAATATAG	TATATTCGAG	ACTCCAGATC	AAAAATCACC	8280
GAGCTCAGG	GAGAAAGAGG	GCCACTAATG	GCCAGAGCCC	TTCAGCCTTC	TCCCACCCCTG	8340
CCTGTACCTT	CAGATGGAAG	CATTTTTTTA	TCATTGTTTC	ACCTTTAGCA	TTTTGACAAT	8400
GAAGTCACAA	ACCTTCAGGC	TCTCACCCAT	AGGAACCCAC	TGGTTGTAAG	AGAAGGATGA	8460
AGCCAGTCTT	TCTTAAGAGG	CACGATTAGA	TGTGTTTATG	GCATCCTCAG	GTGAAACTAT	8520
ATTTATATTG	ACAATATATT	TATATTTCTC	AGGAATAACT	AGAAATAATGA	TTGAGTTTCA	8580
TACTAGGCTA	TTTATCTATC	CTTTATAATA	TTGTTTAAAT	AGAAATGCT	TTCTATCTTC	8640
CAAAATATTG	ATGATTTGTA	AGAGAACACT	TAAACATGGA	TATTCATAAG	CTGAAACTTC	8700
TGGCATTTAT	TGAATGTCAA	GATTTTCTAT	CAGTATACTA	GTTGATTAAC	TGACCACTGA	8760
ACTTGAAGGT	AGTATAAGGT	AGTAGTAAAA	GATACAAATCA	TTGTCTCTTA	ACAGATGGCT	8820
CTTTCTTTTC	ATTAG					8835

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1371 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human  
(F) TISSUE TYPE: placenta

(ix) FEATURE:  
(A) NAME/KEY: intron  
(B) LOCATION: 1..1371  
(C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

```
GTAAGGCTAA TGCCATAGAA CAAATACCAG GTTCA3ATAA ATCTATTCAA TTAGAAAAGA 60
TGTTGTGAGG TGAACATATTA AGTGACTCTT TGTGTCAACCA AATTTCACCTG TAATATTAAT 120
GGCTCTTAAA AAAATAGTGG A2CTCTAGAA ATTAACCACA ACATGTCCAA GGTCTCAGCA 180
CCTTGTCACA CCACGTGTCC T3GCACTTTA ATCAGCAGTA GCTCACTCTC CAGTT3GCAG 240
TAAGTGACAA TCATGAAAAT CCCAGTTTTT ATGGGAAAAAT CCCAGTTTTT ATTGGATTTC 300
CATGGGAAAA ATCCCAGTAC AAAA2CTGGGT GCATTCA3GA AATACAATTT CCCAAAGCAA 360
ATTGGCAAAAT TATGTAAGAG ATTCTCTAAA TTTAGAGTTC CGTGAATTAC ACCATTTTAT 420
GTAAATATGT TTGACAAGTA AAAATTGATT CTTTTTTTTT TTTTCTGTTG CCCAG3CTGG 480
AGTGCA3TGG CACAATCTCT GCTCACTGCA ACCTCCACCT CCTGG3GTTCA AGCAATTCTC 540
CTGCCTCAGC CTTCTGAGTA GCTG33ACTA CAGGTGCATC CCGCCATGCC TGGCTAATTT 600
TTGGGTATTT TTACTAGAGA CAGG3TTTTG GCATGTTGTC CAGGCTGGTC TTGGACTCCT 660
GATCTCAGAT GATCCTCCTG GCTCG3GCTC CCAAAGTGCT GGGATTACAG GCATGAACCA 720
CCACACATGG CCTAAAAATT GATTCTTATG ATTAATCTCC TGTGAACAAT TTGGCTTCAT 780
TTGAAAGTTT GCCTTCATTT GAAAC3TTCA TTTAAAAGCC TGAGCAACAA AGTGAGACCC 840
CATCTCTACA AAAA2CTGCA AAATATCCTG TGGACACCTC CTACCTTCTG TGGAG3CTGA 900
AGCAGGAGGA TCACCTGAGC CTAG3AATTT GA3CCTGCAG TGAGCTATGA TCCCACCCCT 960
ACACTCCAGC CTGCATGACA GTAGACCCTG ACACACACAC ACAAAAAAAA ACCTTCATAA 1020
AAAATTATTA GTTGACTTTT CTTA33TGAC TTTCCGTTTA AGCAATAAAT TTAAAAGTAA 1080
AATCTCTAAT TTTAGAAAAT TTATTTTTAG TTACATATTG AAATTTTTTA ACCCTAGGTT 1140
TAAGTTTTAT GTCTAAATTA CCTGAGAACA CACTAAGTCT GATAAGCTTC ATTTTATGGG 1200
CCTTTTGGAT GATTATATAA TATTCTGATG AAA3CCAAGA CAGACCCTTA AACCATAAAA 1260
ATAGGAGTTC GAGAAAGAGG AGTA3CAAAA GTAAAAGCTA GAATGAGATT GAATTCTGAG 1320
TCGAAATACA AAATTTTACA TATTCTGTTT CTCTCTTTTT CCCCCTCTTA G 1371
```

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 3383 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: human  
(F) TISSUE TYPE: placenta

(ix) FEATURE:  
(A) NAME/KEY: intron  
(B) LOCATION: 1..3383  
(C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

```
GTAAAGTAGA AATGAATTTA TTTTCTTTTG CAACTAAGT ATCTGCTTGA GACACATCTA 60
TCTCA2CATT GTCAGCTGAG GAAAAA3AAA AATGGTTCTC ATGCTACCAA TCTGCTTCTA 120
AAGAAATGTG GACTCAGTAG CACAGCTTTG GAATGAAGAT GATCATAAGA GATACAAAGA 180
AGAA2CTCTA GCAAAAGATG CTTCTCTATG CCTTAAAAAA TTCTCCAGCT CTTAGAATCT 240
ACAAAATAGA CTTTGCTGTG TTCATTG3TC CTAAGATTAG CATGAAG3CA TGGATTCTGT 300
TGTAG333GA G3CTTG3ATA GGAAAAAGGG ATTGAAGCAT TAGAATTGTC CAAAATCA3T 360
AACACCTCCT CTCAGAAATG CTTTG3GAAG AAGCCTGGAA GGTTCCGGGT TGGTGGTGGG 420
GTGG33CAGA AAATTCTGGA AGTAGAG3AG ATAGGAATGG GTG3GGCAAG AAGACCACAT 480
TCAGAG33CA AAAGCTGAAA GAAACCATG3 CATTTATGAT GAATTCAGGG TAATTCAGAA 540
TGGAAGTAGA GTAGGAGTAG GAGACTGGTG AGAG3AGCTA GAGTGATAAA CAGG3TGTAG 600
AGCAAGAGCT TCTCTCACCC CAAGATGTGA AATTGGACT TTATCTT3GA GATAATAGGG 660
TTAATTAAGC ACAATATGTA TTAGCTAGGG TAAAGATTAG TTTGTTGTAA CAAAGACATC 720
```

CAAAGATACA	GTAGCTGAAT	AAGATAGAGA	ATTTTTCTCT	CAAAGAAAGT	CTAAGTAGGC	780
AGCTCAGAAG	TAGTATGGCT	GGAAGCAACC	TGATGATATT	GGGATCCCCA	ACCTTCTTCA	840
GTCTTGTACC	CATCATCCCC	TAGTTGTTGA	TCTCACTCAC	ATAGTTGAAA	ATCATCATAC	900
TTCTTGSGTT	CATATCCCAG	TTATCAAGAA	AGGGTCAAGA	GAAGTCAGGC	TCATTCCCTT	960
CAAAGACTCT	AATTGGAAAT	TAAACACATC	AATCCCCCTC	ATATTCCATT	GACTAGAATT	1020
TAATCACATG	GCCACACCAA	GTGCAAGGAA	ATCTGGAAAA	TATAATCTTT	ATTCCAGSTA	1080
GCCATATGAC	TCTTTAAAAAT	TCAGAAATAA	TATATTTTTA	AAATATCATT	CTGCTTTTGG	1140
TATAAAGAAT	TGATGGTGTG	GGGTGAGGAG	GCCAAAATTA	AGGGTTGAGA	GCCATTATT	1200
TTAGTTATTA	CAAAATATGA	TGGTGTCTAT	AATTAAGSTA	GACATAGGGG	AGTGTGTATG	1260
AGAGCTGTG	AATGATTTT	AGAAACACTT	GAGAGAATCA	ATAGGACATG	ATTTAGGGTT	1320
GGATTTGGAA	AGGAGAAAG	AGTAGAAAA	ATGATGCCTA	CATTTTTTAC	TTAGGCAATT	1380
TGTACCATTC	AATGAAATAG	GGAACACAGG	AGGAAAGACA	GGTTTTGGTG	TATACAAAGA	1440
GGAGATATGA	TGAGGATTTT	CGTTTTGGAT	CTGAGATGTC	TGTGGAACGT	CCTAGTGGAG	1500
ATGTCCACAA	ACTCTTCTAC	ATGTGTTCTG	GATTTCAAGG	CACAGATTTG	GGCTGGAGAT	1560
AGAGATATTG	TATGCTTATA	CATAGAAATG	GCATTTGAAT	CTATAGAAT	AAAAAGACAC	1620
ATCAAGAGAA	ATGTGTAAAG	TGAGAGAGGA	AAAGCCAAGT	ACTGTGCTGG	GGGGAATACC	1680
TACATTTAAA	GGATGCASTA	GAAAGAAGGT	AATAAACAA	AAGAGCAGA	CTAACCAAAA	1740
GGAGAGAGAA	AAAGCAAGA	GAATTCACAC	GATCCACAG	AAGCATTTT	AAGATTGAGG	1800
GGATAGGTGT	TGTGTTGAAT	TTTGCAAGCT	TGAGAATCAA	GGGCCAGAAC	ACAGCTTTTA	1860
GATTTAGCAA	CAAGGAGTTT	GGTGATCTCA	GTGAAAGCAG	CTTATGGTG	AAATGGAGGC	1920
AGAGGACATG	TGCAATGAGT	GAAACAGTGA	ATGGGAAGTG	AAGAAATGAT	ACAGATAATT	1980
CTTGCTAAAA	GCTTGGCTGT	TAAAAGGAGG	AGAGAAACAA	GACTAGCTGC	AAAGTGAGAT	2040
TGGGTTGATG	GAGCAGTTTT	AAATCTCAAA	ATAAGAGCT	TTGTGCTTTT	TTGATTATGA	2100
AAATAATGTG	TTAATTGTAA	CTAATTGAGG	CAATGAAAA	AGATAATAAT	ATGAAAGATA	2160
AAAATATAAA	AACCACCCAG	AAATAATGAT	AGCTACCATT	TTGATACAAT	ATTTCTACAC	2220
TCCTTTCTAT	GTATATATAC	AGACACAGAA	ATGCTTATAT	TTTTATTAAA	AGGGATTGTA	2280
CTATACCTAA	GCTGCTTTTT	CTAGTTAGTG	ATATATATGG	ACATCTCTCC	ATGGCAACGA	2340
GTAATTGCAG	TTATATTAA	TTCATGATAT	TTCAACAATA	GGGCATATCT	TTGCCCTTTT	2400
TATTTAATCA	ATTCTTAATT	GGTGAATGTT	TGTTTTCCAGT	TTGTTGTTGT	TATTAACAAT	2460
GTTCCCATAA	GCATTCCTGT	ACACCAATGT	TCACACATTT	GTCTGATTTT	TTCTTCAGGA	2520
TAAAACCCAG	GAGGTAGAA	TGCTGGGTTG	ATAAAGAGA	AAGGATGATT	GCCAAATTAA	2580
AGCTTCASTA	GAGGGTACAT	GCCGAGCACA	AATGGGATCA	GGCCTAGATA	CCAGAAATGG	2640
CACTTTCTCA	TTTCCCTTTG	GGACAAAAG	GAGAGAGGCA	ATAACTGTGC	TGCCAGAGTT	2700
AAATTTGTAC	GTGAGTATG	AGGAATCAT	TTGCTGAAAA	TGAAACAGAG	GATGATGTTG	2760
TAGAGGCTCT	GAAGAGAGCA	AAGAAATTT	GAAATTTGGG	CTATCAGCTA	TGGAAGAGAG	2820
TGCTGAACAT	GAAACAAAA	GAAGTATTGA	CAATTTGSTAT	GCTTGTAATG	GCACCGATTT	2880
GAACGCTTGT	GCCATTGTTT	ACCAGCAGCA	CTCAGCAGCC	AAGTTTGGAG	TTTTGTAGCA	2940
GAAAGACAAA	TAAATTAGGG	ATTTAATATC	CTGCCCCAAT	GCTAGACAAA	ATGAACTCTG	3000
AGATCCAGCT	GCACAGGGAA	GGAAGGGAAG	ACGGGAAGAG	GTTAGATAGG	AAATACAAGA	3060
GTCAGGAGAC	TGGAAGATGT	TGTGATATTT	AAAGACACAT	AGAGTTGGAG	TAAAAGTGTA	3120
AGAAAACCTAG	AAGGGTAAGA	GACCGGTCAG	AAAGTAGGCT	ATTTGAAGTT	AACACTTCAG	3180
AGSCAGAGTA	GTTCTGAATG	GTAACAAGAA	ATTGAGTGTG	CCTTTGAGAG	TAGGTTAAAA	3240
AAACAATAGGC	AACCTTTATTG	TAGCTACTTC	TGGAACAGAA	GATTGTCATT	AATAGTTTTA	3300
GAAAACTAAA	ATATATAGCA	TACTTATTTG	TCAATTAACA	AAGAACTAT	GTATTTTTTA	3360
ATGAGATTTA	ATGTTTATTG	TAG				3383

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11464 base pairs
- (E) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
- (F) TISSUE TYPE: placenta

(ix) FEATURE:

- (A) NAME/KEY: 5'UTR
- (B) LOCATION: 1..3
- (C) IDENTIFICATION METHODS: E
- (A) NAME/KEY: leader peptide
- (B) LOCATION: 4..82
- (C) IDENTIFICATION METHODS: S

(A) NAME/KEY: intron  
 (B) LOCATION: 83..1453  
 (C) IDENTIFICATION METHODS: E  
 (A) NAME/KEY: leader peptide  
 (B) LOCATION: 1454..1465  
 (C) IDENTIFICATION METHODS: S  
 (A) NAME/KEY: intron  
 (B) LOCATION: 1466..4848  
 (C) IDENTIFICATION METHODS: E  
 (A) NAME/KEY: leader peptide  
 (B) LOCATION: 4849..4865  
 (C) IDENTIFICATION METHODS: S  
 (A) NAME/KEY: mat peptide  
 (B) LOCATION: 4866..4983  
 (C) IDENTIFICATION METHODS: S  
 (A) NAME/KEY: intron  
 (B) LOCATION: 4984..6317  
 (C) IDENTIFICATION METHODS: E  
 (A) NAME/KEY: mat peptide  
 (B) LOCATION: 6318..6451  
 (C) IDENTIFICATION METHODS: S  
 (A) NAME/KEY: intron  
 (B) LOCATION: 6452..11224  
 (C) IDENTIFICATION METHODS: E  
 (A) NAME/KEY: mat peptide  
 (B) LOCATION: 11225..11443  
 (C) IDENTIFICATION METHODS: S  
 (A) NAME/KEY: 3'UTR  
 (B) LOCATION: 11444..11464  
 (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

AAG ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA	48
Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala	
-35 -30 -25	
ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G GTAAGG CTAATGCCAT	98
Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala	
-20 -15 -10	
AGAACAAATA CCAGSTTCAG ATAAATCTAT TCAATTAGAA AAGATGTTGT GAGGTGAACT	158
ATTAAGTGAC TCTTTGTGTC ACCAAATTTT ACTGTAATAT TAATGGCTCT TAAAAAATA	218
GTGGACCTCT AGAAATTAAC CACAACATGT CCAAGGTCTC AGCACCTTGT CACACCAAGT	278
GTCCTGGCAC TTTAATCAGC AGTAGCTCAC TCTCCAGTTG GCAGTAAGTG CACATCATGA	338
AAATCCCAST TTTGATGGGA AAATCCCAST TTTTCATTGGA TTTCCATGGG AAAAAATCCCA	398
GTACAAAACCT GGGTSCATTC ABGAAATACA ATTTCCCAAA GCAAATTGSC AAATTATGTA	458
AGAGATTCTC TAAATTTAGA GTTCCSTGAA TTACACCATT TTATGTAAAT ATGTTTGACA	518
AGTAAAAATT GATTCTTTTT TTTTTTTTTT GTTGCCCAAG CTGGAGTGCA GTGGCACAAT	578
CTCTGCTCAC TGCAACCTCC ACCTCCTGGG TTCAAGCAAT TCTCCTGCCT CAGCCTTCTG	638
AGTAGCTGGG ACTACAGGTG CATCCCGCCA TGCCCTGGCTA ATTTTGGGGT ATTTTACTA	698
GAGACAGEST TTTGGCATST TGTCCAGGCT GGTCTTGGAC TCCTGATCTC AGATGATCCT	758
CCTGCTCGG GTTCCCAAAG TGTCTGGATT ACAGGCATGA ACCACCACAT ATGGCCATAA	818
AATTGATTCT TATGATTAAT CTCTCTGAA CAATTTGGCT TCATTTGAAA GTTTGCCCTC	878
ATTTGAAACC TTCATTTAAA AGCTGAGCA ACAAGTGAAG ACCCCATCTC TACAAAAAAC	938
TGCAAAATAT CTTCTGGACA CTTCTACCT TCTGTGGAGG CTGAAAGCAGG AGGATCACTT	998
GAGCCTAGGA ATTTGAGCCT GAGTGGAGCT ATGATCCAC CCCTACACTC CAGCCTGCAT	1058
GACAGTAGAC CTTGACACAC ACACACAAA AAAAACCTTC ATAAAAAATT ATTAGTTGAC	1118
TTTTCTTAGG TGACTTTCCG TTTAAGCAAT AAATTTAAAA GTAAAAATCTC TAATTTTAGA	1178
AAATTTATTT TTAGTTACAT ATTGAAATTT TTAACCCCTA GGTTTAAGTT TTATGTCTAA	1238
ATTACCTGAG AACACACTAA GTCTGATAAG CTTCATTTTA TGGGCCTTTT GGATGATTAT	1298
ATAATATTCT GATGAAAGCC AAGACAGACC CTTAAACCAT AAAAATAGGA GTTCGAGAAA	1358
GAGGAGTAGC AAAAGTAAAA GCTAGAATGA GATTGAATTC TGAGTCGAAA TACAAAATTT	1418
TACATATTCT GTTTCTCTCT TTTTCCCCCT CTTAG CT GAA GAT GAT G GTAAA	1470
Ala Glu Asp Asp Glu	
-10	
GTAGAAATGA ATTTATTTTT CTTTGCAAA TAAGTATCTG CTTGAGACAC ATCTATCTCA	1530
CCATTGTCAG CTGAGGAAAA AAAAAATG3 TTCTCATGCT ACCAATCTGC CTTCAAAGAA	1590
ATGTGGACTC AGTAGCACAG CTTTGGAATG AAGATGATCA TAAGAGATAC AAAGAAGAAC	1650

CTCTA3CAAA	AGATGCTTCT	CTATGCTTCA	AAAAATTCTC	CAGCTCTTAT	AATCTACAAA	1710
ATAGACTTTT	CCTGTTTCAT	TGGTCTTAAG	ATTAGCATGA	AGCCATGGAT	TCTGTTGTAG	1770
GGGGAGCGTT	GCATAGGAAA	AAGGGATTGA	AGCATTAGAA	TTGTCCAAAA	TCAATAACAC	1830
CTCCTCTCA3	AAATGCTTTG	G3AAGAAGCC	TGGAAGGTTT	CGGTTTGGTG	GTGGGGTGGG	1890
GCAGAAAATT	CTG3AAGTAG	AG3AATAGG	AATGGGTGGG	GCAAGAAGAC	CACATTCAGA	1950
G3CCAAAAGC	TGAAA3AAAC	CATGGCATTT	AT3ATGAATT	CAG3GTAATT	CAGAATGGAA	2010
GTAG3ATAG3	A3TA33AGAC	TG3TGAGAG3	AG3TAGAGT3	ATAAACAG33	TGTAGAGTAA	2070
GAG3TTCTCT	CACCCCAAGA	T3TGAAATTT	GGACTTTTAT	TT33AGATAA	TAGGGTTAAT	2130
TAAGCACAAT	ATGTATTAGC	TAG33TAAAG	ATTAGTTTGT	TGTAACAAA3	ACATCCAAA3	2190
ATACAGTAGC	TGAATAAGAT	AGAGAAATTT	TCTCTCAAA3	AAA3TCTAA3	TAGGCAGCTC	2250
AGAAGTAGTA	TGGCTGGAAG	CAACCTGATG	ATATTG33AC	CCCAACCTTT	CTTCAGTCTT	2310
GTACCCATCA	TCCCCTAGTT	GTTGATCTCA	CTCACATAGT	TGAAAATCCT	CATACTTCTT	2370
G33TTTCATAT	CCCATTTATC	AAG3AAGGGT	CAAGAGAAGT	CA33CTCATT	CCTTTCAAA3	2430
ACTCTAATTG	GAAGTTAAAC	ACATCAATCC	CCCTCATATT	CCATTGACTA	GAATTTAATC	2490
ACATG3CCAC	ACCAAGTGCA	AG3AAATCTG	GAAAATATAA	TCTTTATTCC	AG3TAGCTAT	2550
ATGACTCTTT	AAAAATTCAGA	AATAATATAT	TTTTAAATAA	TCATTCTG3C	TTTGGTATAA	2610
AGAATT3ATG	GTGTGGGGTG	AGGAGGCCAA	AATTAAGGGT	TGA3AGCCTA	TTATTTTAGT	2670
TATTACAAGA	AATGATGGTG	TCATGAATTA	AG3TAGACAT	AGGGGAGTGC	TGATGAG3AG	2730
CTGTGAATG3	ATTTTAGAAA	CATTGAGAG	AATCAATAG3	ACATGATTTA	GG3TTGGATT	2790
TGGAAA33AG	AAGAAAGTAG	AAA3ATGAT	GCCTACATTT	TTCACTTAG3	CAATTTGTAC	2850
CATTCA3TGA	AATAG33AAC	ACAGGAGGAA	GATCAGGTTT	TGGTGTATAC	AAAGAGGAGG	2910
ATGGAT3ACG	CATTTCGTTT	TGGATCTGAG	ATGTCTGTGG	AACGTCCTAG	TGGAGAT3TC	2970
CACAAA3TCT	TCTACATGTG	GTTCTGAGTT	CAGGACACAG	ATTTGGGCTG	GAGATAGAGA	3030
TATTGTAG3C	TTATACATAG	AAATGGCATT	TGAATCTATA	GAGATAAAAA	GACACATCAG	3090
T3GAAATGTC	TAAAGTGAGA	GAG3AAAAGC	CAAGTACTGT	GCTGGGGG3A	ATACCTACAT	3150
TTAAAG3ATG	CAGTGAAGAG	AGCTAATAAA	ACAACAGAG3	GCAGACTAAC	CAAAAGGGGA	3210
GAAGAAAAAC	CAAGAGAATT	CCATGGACTC	CCAG3AGAG3	ATTTCAAGAT	TGAGGGGATA	3270
G3TGTT3TGT	TGAATTTTGC	AGCCTTGAGA	ATCAAG3GOC	AGAACACAGC	TTTTAGATTT	3330
AGCAACAAG3	AGTTTGGTGA	TCTCAGTGAA	AGCAGCTTGA	TGGTGAAATG	GAGGCAGAGG	3390
CAGATT3CAA	TGAGT3AAAC	AGT3AATGG3	AGTGAAGAA	ATGATACAG3	TAATTTCTGC	3450
TAAAG3CTTG	GCTGTTAAAA	GGAG3AGAGA	AACAAGACTA	GCTGCAAA3T	GAGATTGGGT	3510
TGATGGAG3A	GTTTTAAATC	TCAAAATAAA	GAGCTTTGTG	CTTTTTTTGAT	TATGAAAATA	3570
ATGTGTTAAT	TGTAACATAA	TGAG3CAAT3	AAAAA3ATA	ATAATATGAA	AGATAAAAAAT	3630
ATAAAAAACA	CCCAGAAATA	ATGATAGCTA	CCATTTTGAT	ACAATATTTT	TACACTCTCT	3690
TCTAT3TATA	TATACAGACA	CAGAAATGCT	TATATTTTAA	TTAAAAGG3A	TTGTACTATA	3750
CCTAAGCT3C	TTTTTCTAGT	TAGT3ATATA	TATGGACATC	TCTCCAT33C	AACGAGTAAT	3810
T3CAGTTATA	TTAAGTTTAT	GATATTTTAC	AATAAG3GCA	TATCTTTG3C	CTTTTTATTT	3870
AATCAATTTT	TAATTT3GT3A	ATGTTTGTTT	CCA3TTTGTT	GTT3TTATTA	ACAATGTTCC	3930
CATAAGCATT	CCTGTACACC	AATGTTTACA	CATTTTGCTG	ATTTTTTTCT	CAGGATA3AA	3990
CCAG3AG3T	AGAATTTGCT	G3TT3ATAGA	AG3AAGGGA	TGATTGCTAA	ATTAAGCTTT	4050
CAGTAGAG3G	TACAT3CC3A	GCACAAT3G	GATCAGCCCT	AGATACC3A	AATG3CACTT	4110
TCTCATTTCC	CCTT33GACA	AAAG3GAGAG	AG3CAATAAC	TGTGCTG3CA	GAGTTAAATTT	4170
T3TAC3T33A	GTAG3AG3AA	ATCATTTTGT	GAAAATGAAA	ACAGAGATGA	TGTTGTA3AG	4230
GTCT3A33A	GAGCAAA3AA	AATTTGAAAT	TGCGGCTATC	AGCTATG3AA	GAGAGTGCTG	4290
AATG33AAA	CAAAAGAG3T	ATT3ACAATT	G3TATGCTTG	TAATGG3ACC	GATTTGAA3G	4350
CTTGT3CCAT	TGTTCA3CAG	CAGCACTCAG	CAGCCAA3TT	TGAGTTTTTG	TAGCAGAAAG	4410
ACAAATAA3T	TAG3GATTTA	ATATCTG33C	CAAATG3TAG	ACAAAATGAA	CTCTGAGATC	4470
CAGCTGCACA	GGGAAG3AA3	GGAG3AC33G	AAG33TTAG	ATAG3AAATA	CAAGAGTCAG	4530
GAGACTG3AA	GATGTT3T3A	TATTTAA3AA	CACATAGAGT	TG3AGTAAAA	GTGTAAG3AA	4590
ACTAGAG33G	TAGAGAG33G	GTCA3AAAT3	AG3CTATTTG	AGGTTAACA3	TTCAGAG3CA	4650
GAGTAGTTCT	GAAT33TAA3	AAGAAATTT3	GT3T3CCTTT	CAGAGTAG3T	TAAAAACAA	4710
TAG3CAACTT	TATTTGAT3T	ACTTTCTG3A	CAGAA3ATTT	TCATTAATA3	TTTTAGAAAA	4770
CTAAATATA	TAGCATACTT	ATTGTCATAT	TAACAAAGAA	ACTATGTATT	TTTAAT3AG	4830
ATTTAAT3T	TATTTGTAG	AA AAC CTG	GAA TCA GAT	TAC TTT GGC AAG CTT		4880
		Glu Asn Leu	Glu Ser Asp	Tyr Phe Gly Lys Leu		
		-5		1	5	
GAA TCT AAA TTA TCA GTC ATA AGA AAT TTG AAT	GAC CAA GTT CTC TTC					4928
Glu Ser Lys Leu Ser Val Ile Arg Asn Leu Asn	Asp Gln Val Leu Phe					
		10	15	20		
ATT GAC CAA G3A AAT CG3 CCF CTA TTT GAA GAT	ATG ACT GAT TCT GAC					4976
Ile Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp	Met Thr Asp Ser Asp					
		25	30	35		
TGT AGA G GTATTTTTT TTAATTTGCA AACATAGAAA	TGACTAGCTA CTTCTTCCCA					5032
Cys Arg Asp						
		40				
TTCTGTTTTA CTGCTTACAT TGTTCCGTGC TAGTCCCAAT	CCTCAGATGA AAAGTCACAG					5092
GAGTGACAAT AATTTCACTT ACAGGAAACT TTATAAG3CA	TCCAC3TTTT TTAGTTGGGG					5152

TAAAAAATTG	GATACAATAA	GATATTGCTA	GGGGTATATG	CTCTCTGAGG	CTGCCTTTGA	5212
ATCAACCAAT	CCTTTATTGT	GATTGCATTA	ACTGTTTTAA	ACCTCTATAG	TTGGATGCTT	5272
AATCCCTGCT	TGTTACAGCT	GAAAATGCTG	ATAGTTTTAC	AGGTGTGTGT	GCATCTATCT	5332
GTAATCCTAG	CTACTTG3GA	GGCTCAAGCA	GGAGGATTG	TTGAGGCTAG	GACTTTT3AG	5392
CTGTAGTACA	CTGTGATGCT	ACCTGTBAAT	AGGCACTGCA	CTCCAGGCTG	GGTGATATAC	5452
AGACCTTTGT	TCTAAAAATT	AAAAAAAATA	AAAAAAAATA	CTT3AGAAAG	GAAATTGATC	5512
AACTCTACTG	TGCCTTCCAA	AAATGSAATT	CAAAATATCA	AAATTA33CT	GAGTTGAAAG	5572
AGTGAATGTT	CATTCTTTAA	AAATACTGAA	TACTTAACCT	AACATATATT	TTAAATATTT	5632
TATTTAGCAT	TTAAAAGTTA	AAAACAATCT	TTTAGAATTC	ATATCTTTAA	AATACTCAAA	5692
AAAGTTGCA	CGTGTGTGTT	GTAATACACA	TTAAACTGTG	GGGTTGTTTG	TTTGTGTTG	5752
ATGCAGTTTT	ACTCTGTGAC	CCAGGCTGAA	GTGCAGTGC	GTGCAGTGGT	GTGATCTCG	5812
CTCACTACAA	CCTCCACCTC	CCAGGTTCAA	GGGATTCTCA	TGCTCTA3TC	TCCCGAGTAG	5872
GTGGGATTA	AGGCATGAC	CACTTACAC	CGGCTAATTT	TTTATTTTTT	AGTAGAGCTG	5932
GGGTTTCACT	ATGTTGGCCA	GGCTGCTCT	AAACCCCTAA	CCTCAAGTGA	TCTGCTGCT	5992
TCAAGCTCCC	AAACAAACAA	ACAACCCAC	AGTTTAAAT	GTGTTACAAC	ACACATGCTG	6052
CAACTTTTAT	GAGTATTTTA	ATGATATAGA	TTATAAAAG	TTGTTTTTAA	CTTTTAAATG	6112
CTGGGATTA	AGGCATGAGC	CACTGTGCTA	GGCTGAACT	GTGTTTTTAA	AAATGCTCTG	6172
CCAGCTGTAC	ATAGTCTCCT	GCAGAGTGG	CAAGTCTCAA	AGTGGGAACA	GGTGTATTAA	6232
GACTATCTCT	TTGGTTAAAT	TTCCGCAAT	GTTCCTGTGC	AA3AATTCTT	CTAAGTAGAG	6292
TTCTCATTTA	TTATATTTAT	TTCA	AT AAT GCA CCC	CGG ACC ATA TTT ATT		6343
		Asp	Asn Ala Pro	Arg Thr Ile Phe Ile		
		40		45		
ATA AGT ATG TAT AAA GAT AGC CAG CCT AGA G3T	ATG GCT GTA ACT ATC	6391				
Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly	Met Ala Val Thr Ile					
50	55	60				
TCT GTG AAG TGT GAG AAA ATT TCA ACT CTC TCC	TGT GAG AAC AAA ATT	6439				
Ser Val Lys Cys Glu Lys Ile Ser Thr Leu Ser	Cys Glu Asn Lys Ile					
65	70	75				
ATT TCC TTT AAG GTAAG ACTGAGCCTT ACTTTGTCTT	CAATCATGTT AATATATCA	6495				
Ile Ser Phe Lys						
ATATAATTAG AAATATAACA TTATTTCTAA TGTTAATATA	AGTAATGTAA TTAGAAACT	6555				
CAAAATATCCT CAGACCAACC TTTTGTCTAG AACAGAAATA	ACAAGAAGCA GAGAACCAT	6615				
AAAGTGAATA CTTACTAAAA ATTATCAAA	TCTTTACCTA TTGTGATAAT	6675				
CTGAGCCTGT CACAGGGGAA GAGGAGATAC	AACACTTGTT TTATGACCTG	6735				
ACAATCAGTC TTTATACAAA TAATAATGTA	GAATACATAT GTGAGTTATA	6795				
TAAATGATGA CTTTCCAGAA TGAGTTCTGT	TATGAGAAAT ATCCTTCTAT	6855				
ATTTCTACAC CTTTGTAAAT TATGATAATA	TTTTAATCCC TAGTTGTTTT	6915				
CTTAGCTTAA GTCTTAGACA CAAGCTTCAG	CTTCCAGTTG ATGATATGTTA	6975				
TAATCTAATT GAATAAAGT TATGAGATCA	GCTGTAAAG TAATGCTATA	7035				
AGCCAGGTAT AAGTATTTT TGGCTCTCTA	TTTTCTCTTA TTATTCTCTA	7095				
CTATTATTTT TCTCTATTTT CTCCATTATT	GTTAGATAAA CCACAAATTA	7155				
AGACTGAGGC AGTAAGASTA GGCAGGATG	CTTACAAATF GGCATGCTT	7215				
TTCCATGTCA TGAAGACTCT TTTTGAATG	AGATTTGCCA ATAAATATCC	7275				
CCACCCACTC CCACCTGAAA GACAGTTAG	ATATGACCTT AGTGAAGSTA	7335				
ACTTGGTAGG GAGAAAAAG CCACCTCTAA	ATATATCCA AGTAA3AACA	7395				
AACAGATACA GCGCCCTGAC AAATCCCTCA	GCTATCTTCC TCCAAACAGA	7455				
TTGAGGTGAC AATTTGAGT CCCCATTCTA	GACCTGACAG GCAGCTTAGT	7515				
GCATAAGAGG CCTGCGATG	AAAG3TAGGG TGGAAAGGTT	7575				
AACATAATTA GAAGGGAGG AGAT33CAAA	GCTCAGCTA TGTG3BATAG	7635				
AGCTGAGAG GAGATTGAG AAAT33GAT	AAATCTGAGC CTACAGGTGG	7695				
AGGGAGACTG GTGAAATGT TAAGAAAGATG	GAATCAATGC TTG3CACTTA	7755				
GGCAAAATCCA TATTTGGGG AGCT3AAAT	TTATTTCAAT TTGATGGCCC	7815				
AAAGAATGTG GCTGGG3GTG GTGG33CAAA	CCTGTAATCC CAGCACTTTG	7875				
GGGGGGGAT CACCTGAAGT CAG3ASTTCA	AGA3CAGCTT GACCAACAT3	7935				
ATCTCTACTA AAAATACAAA ATTA3CTGG	CGTGGTGGCA TAT3CTGTA	7995				
CTGGGGAGGC TGAG3CA3A GAATCTTTG	AACCG3GAG GCAGAGBT3	8055				
AGAT3CTGCC ATTG3ACTCC AGTCT33CA	ACAAGAGCAA AACTG3TCT	8115				
AAAAAAAAG TGAAATTAAC CAAG33CAT	AGCTTAATAA TTTAATACT3	8175				
GGGGGGGGG TGGCTGGAAG AGATCT3TGT	AAATGAGGGA ATCTGACATT	8235				
CAGCATCATA GCAAACTCTC TTCTG3AAG	AATTCATAA ATATTAGTTG	8295				
GAGAGTGAGG GGTGGAGTAG GACCA3TTT	AGCTTTGTC TTTAATCCCT	8355				
CTAATAAG3A TCTTAG3CA3T GGTATAAAA	GT33CTAG3 TTCTAGATAA	8415				
CAG33CAG3C ACA3T33CTC AT33CTATAA	TCCAG3CACT TT33AG33C	8475				
TCTCACTTGA GATCAG3AGT TCAAGA3CAG	CCTGG3CAGC AT33CGATAC	8535				
CTAAAAAAA TACAAAAAT AGCCAGGCAT	G3T33CAT3C AACTGTAATC	8595				
GTGAG3CCTGA GGCAGAAAGAA TGGCTTGA	CCAG33T3G TAG3CTGCAG	8655				
TGCACCACT GCACTCCAGC CTGG33GACA	GAATGAGACT TTGTCTCAAA	8715				

GATACAACAG	GCTACCCTTA	TGTGCTCAAC	TTTCACTGTT	GATTACTAGC	TATAAAGTCC	8776
TATAAAGTTC	TTTGGTCAAG	AACCTTGACA	ACACTAAGAG	GGATTTGCTT	TGAGAGGTTA	8836
CTGTCAAGT	CTGTTTCATA	TATATACATA	TACATGTATA	TATGTATCTA	TATCCAGGCT	8896
TGGCCAGGTT	TCCCTCAGAC	TTTCCASTGC	ACTTGGGAGA	TGTTAGGTCA	ATATCAACTT	8956
TCCCTG3ATT	CAGATTCAAC	CCCTTCTGAT	GTAAGAAAAA	AAAAAAAAAA	GAAAGAAATC	9016
CCTTTCCCTT	TGGAGCACTC	AAGTTTCAAC	AGTTGGG3CT	TTCCAA3TTG	GGGGTTCTCC	9076
AAG3TCATTG	GGATTGCTTT	CACATCCATT	TGCTATGTAC	CTTCCCTATG	ATGGCTGGGA	9136
GTG3TCAACA	TCAAAACTAG	GAAAGCTACT	GCCCAAG3AT	GTCCCTTACCT	CTATTCT3AA	9196
ATGTGCAATA	A3TGTGATTA	AAGAGATTGC	CTGTTCTAC	TATCCACACT	CTCGCTTTCA	9256
ACTGTAACCT	TCTTTTTTTC	TTTTTTCTTT	TTTTTCTTTT	TTTTTGAAAC	GGAGTCTGCT	9316
TCTGTCC3CC	AG3CTAGAGT	GCAGTGG3AC	GATCTCABCT	CACTGCAA3C	TCTG3CTCC	9376
GGGTTCAAGC	CATTCTCCTG	CCTCACCTTC	CCAAGCA3CT	GG3ACTACAG	GCG3CTG3CA	9436
CCATGCCCCA	CTAATTTTTT	GTATTTTTAG	TAGAGACGGG	GTTTCACCGT	GTTAGCCAGG	9496
ATG3TCTG3A	TCTCCT3AAC	TT3TGATCCG	CCCGCCTCAG	CCTCCCAAAG	TGCTGGGATT	9556
ACAGGCT3GA	GCCATCGCAC	CCGCTCAAC	TGTAACTTTC	TATACTGGTT	CATCTTCCCC	9616
TGTAAT3TTA	CTAGAGCTTT	TGAAGTTTTG	GCTATGGATT	ATTTCTCATT	TATACATTAG	9676
ATTTCA3ATT	A3TTCCAAAT	TGATGCCAC	AGCTTAG33T	CTCTTCTAA	ATTGTATATT	9736
GTAGACAGCT	GCAGAAGTGG	GTGCCAATAG	G33AACTAGT	TTATACTTTC	ATCAACTTAG	9796
GACCCACACT	TGTTGATAAA	GAACAAAGST	CAAGAGTTAT	GACTACTGAT	TCCACAAGT	9856
ATTGAGAA3T	TGGAGATAAT	CCCGTGACCT	CTG3CATCCA	GAGTCTTTCA	GGCATCTTTG	9916
AAGGAT3AAG	AAATGCTATT	TAAATTTTGG	AG3TTTCTCT	ATCAGTGCTT	AGGATCATGG	9976
GAATCT3TGC	TGCCATGAGG	CCAAAATTAA	GTCCAAAACA	TCTACTGGTT	CCAGGATTAA	10036
CATGGA33AA	CCTTAGGTGG	TGCCACATG	TTCTGATCCA	TCCTGCAAAA	TAGACATGCT	10096
GCACTAACAG	GAAAAGTGCA	G3CAGCACTA	CCAGTTGGAT	AACCTGCAAG	ATTATAGTTT	10156
CAAGTAATCT	AACCATTTCT	CACAAGGCC	TATTCTGTGA	CTGAAACATA	CAAGAACTG	10216
CAATTG3CCT	TCTAAGGCAG	G3CCCAGCCA	A33AGACCAT	ATTCA3GACA	GAAATTCAAG	10276
ACTACTATG3	AACTGGAGTG	CTTGGCAGGG	AA3ACAGAST	CAAGGACTGC	CAACTGAGCC	10336
AATACAC3AG	GCTTACACAG	GAACCCAGGG	CCTAGCCCTA	CAACAATTAT	TGGGTCTATT	10396
CACTGTAA3T	TTTAATTTCA	GGCTCCACTG	AA3AGTAAG	CTAAGATTCC	TGGCACTTTC	10456
TGCTCTCTCT	ACAGTTG3CT	CAGAAAT3AG	AACTGGTCA	GCCAG3CATG	GTGGCTTACA	10516
CCTGGAATCC	CAG3ACTTTG	GGAG3CCGAA	GT33AGGG3T	CACTTGAGGC	CAGGAGTTCA	10576
GGACCA3CTT	AG3CAACAAA	GTGAGATACC	CCCTGACCCC	TTCTCTACAA	AAATAAATTT	10636
TAAAAATTAG	CCAAATGTGG	TGGT3TATAC	TTACAGTCCC	AGCTACTCAG	GAGGCTGAGG	10696
CAGG3G3ATT	GCTTGAGCCC	A3GAATTCAA	G3CTSCAGTG	AGCTATGATT	TCACCACTGC	10756
ACTTCT33CT	GG3CAACAGA	GCGAGACCTC	GTCTCAAAGC	AAAA33AAA	AGAACTAGA	10816
ACTAGCTTAA	GTTTGTGGGA	GGAG3TCATC	AT3CTCTTTA	GCCGT3AATG	GTTATTATAG	10876
AGGACAGAAA	TTGACATTAG	CCCAAAAAGC	TT3TGGTCTT	TGCTG3AACT	CTACTTAATC	10936
TTGAGCTAAT	GT33ACACCA	CTCAAT33GA	GAG3AGAGAA	GTAAGCTGTT	TGATGTATAG	10996
G3GAA3ACTA	GAG3CCT3GA	ACTGAATATG	CATCCCATGA	CAGGGAGAAT	AGGAGATTCC	11056
GAGTTAAGAA	GGAGAG3AGG	TCAGTACTGC	T3TTCAGAGA	TTTTTTTAT	GTAACCTCTG	11116
AGAAGCA3AA	CTACTTTTGT	TCTGTTT3GT	AATATACTTC	AAACCAAACT	TCATATATTC	11176
AAATT3TTCA	TGTCTTGAAA	TAATTAG3TA	AT3TTTTTTT	CTCTATAG	GAA ATG AAT	11233
					Glu Met Asn	
					85	
CCT CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG						11281
Pro Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Glu						
	90		95		100	
AGA AGT GTC CCA GGA CAT GAT AAT AAG AAG CAA TTT GAA TCT TCA TCA						11329
Arg Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu Ser Ser Ser						
	105		110		115	
TAC GAA GGA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA						11377
Tyr Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys						
	120		125		130	135
CTC ATT TTG AAA AAA GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC						11425
Leu Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe						
	140		145		150	
ACT GTT CAA AAC GAA GAT TAGGTATTAA AATTTCTATGC C						11464
Thr Val Gln Asn Glu Asp						
	155					

(2) INFORMATION FOR SEQ ID NO:14:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28994 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(F) TISSUE TYPE: placenta

(ix) FEATURE:

(A) NAME/KEY: 5'UTR

(B) LOCATION: 1..15606

(C) IDENTIFICATION METHODS: E

(A) NAME/KEY: leader peptide

(B) LOCATION: 15607..15685

(C) IDENTIFICATION METHODS: S

(A) NAME/KEY: intron

(B) LOCATION: 15686..17056

(C) IDENTIFICATION METHODS: E

(A) NAME/KEY: leader peptide

(B) LOCATION: 17057..17068

(C) IDENTIFICATION METHODS: S

(A) NAME/KEY: intron

(B) LOCATION: 17069..20451

(C) IDENTIFICATION METHODS: E

(A) NAME/KEY: leader peptide

(B) LOCATION: 20452..20468

(C) IDENTIFICATION METHODS: S

(A) NAME/KEY: mat peptide

(B) LOCATION: 20469..20586

(C) IDENTIFICATION METHODS: S

(A) NAME/KEY: intron

(B) LOCATION: 20587..21920

(C) IDENTIFICATION METHODS: E

(A) NAME/KEY: mat peptide

(B) LOCATION: 21921..22054

(C) IDENTIFICATION METHODS: S

(A) NAME/KEY: intron

(B) LOCATION: 22055..26827

(C) IDENTIFICATION METHODS: E

(A) NAME/KEY: mat peptide

(B) LOCATION: 26828..27046

(C) IDENTIFICATION METHODS: S

(A) NAME/KEY: 3'UTR

(B) LOCATION: 27047..28994

(C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

ACTTGCCCTTA	AAAGCTTTTGC	ATAGGTAGAC	AACATTAGAT	TAATTTTCCTT	GCTCACATCT	60
GTTCAAGAAA	AATCATTTTAA	GTTATAAAAT	ATAACAAACC	TTCTGCATTA	TAAGACTGAT	120
GTTTAGAAAT	ATAAACATTT	TATACATCAC	CATTTAAATC	TTTCTCCAAG	GCTTCATCTT	180
TATAAAATAG	TCCGAAAATT	TCAGABAAAG	ATGAATCTGA	TTTTCCAAGA	GAGGACAGCT	240
GTGGACTATC	TGGTACTGSA	GACTAAATAA	AGAAAGCAGG	TACAGTCAAT	AAGATCTTCA	300
GGACATATAC	ATTTTGTTTA	TTAAGAAAAA	GCAAATAAAA	CATTTTTTCAG	AAAAAGGCAA	360
ACATGCTAGA	AAGCATATGA	CTTATTCATT	TGAGTTTTTA	TTATTAAGGA	AATTTACAGG	420
CCCAAGAAAC	ACCTTGCTCA	ATATATTAAA	TTTTATTTTG	GTTTTCAACT	AGACTTTGCT	480
TTTCATTTGT	TTGTTTTTBT	GACAAATTC	CGCTCTGTCA	CCTAGGCCAA	AGTGTAGTGA	540
CACAATCTTA	GCTTACTGTA	GCTCTCTAGA	TTCAAGTGAT	CCTCCTGTCT	CAGACTCCTG	600
AGTAGCTAGG	ACTACAGGAA	CATTACACCA	TGCCCAGCTA	ATTTTGTTTT	GTTTTGTTTT	660
GTTTTTCAGG	ACAATGTATT	GCAGGTTTSC	CCAGGCTGAT	CTGAAACTCT	TAGCTCTCAA	720
CGATACTCCT	GCTTCAGCCT	CCCAAGGCAC	TAGGATTACA	GACATGAGCC	AATGCGCCCA	780
GCCTTAAATT	AGACTTTAAA	TGTGTTTTTA	AACTCCTGTT	GAAAAAGCGT	CTGSTATCTT	840
GAACCAAGTAG	ATGTTTTTCAT	AGCAATGAA	CTAAACTGTA	ATTTAGACAG	TAGCCAAATG	900
CTTGTGAAAT	TTTGCTAAAT	AATATAATCT	TCAAGGGAGC	AAATCATGTC	CCAAATGCAA	960
AAGATCAACT	GCTGSGGSCA	GTAGTAAAA	ACAGGATACT	GTGCTCTTTA	AAAGGTCAGT	1020
AACTATAGTA	CCTAGTTATC	TTACTTATCA	CAGCAAAATA	ATTACATAAA	ATCCTATGGA	1080
TCATAAAGGC	ACAGACTCAC	TTCTGTCTCT	AGATCTCAAG	CTACCAAAAA	GAAATCTCCC	1140
AATAGTTTCT	TGGAGGCTCA	TACTTAGTGA	AAAAGCAGCT	GGAATCAACA	TAGTTCTCTC	1200
TATGTTGTAG	GACAATCCTA	GCTCTGSGCA	TACGAATACA	TTAAATCCCA	CTTATCTATA	1260

GAGCTTTCTT	AAAGGGAAGA	AATTTGAGTA	GTATGTAAAA	CAGAATAAAA	GATTAAGGCT	1320
CCATAGGCAT	ACAGCTTACC	TCCAATTCTC	TTGGCCTCTT	GCAATTTCTA	TTATCAGGCT	1380
TTACAAGGTG	ATTTGCTATC	ATATTCGAA	GGCACCAGCT	ACAAAGCTTA	GAACAATGCC	1440
AGATTTAGGT	ACAAACTCCA	TGCTACAAGC	TCTCTGGAAT	CCTTCCCTGT	TTCCCACTCC	1500
TACTGCTGAT	GTTAATTTAG	ACTGTCTATTA	TCTSTCACTT	TCCTAAACTC	AATTTCTTCC	1560
TCCTCTAAAT	CATTCTATCA	ACTGCTATTT	GGGTAATCTT	TCAAAACCTT	GATTACTGCA	1620
TTCCCTTAAAC	TCAAAACCTT	TCATTGTTCC	AGAATAAGTT	GAAATTCAT	GATATGGGCT	1680
TCAGGCTOCT	GTATTATCTG	GTGCAAGGCT	ACTASTCCCA	TCATTTTCAA	CTATCTCTCT	1740
CTATGTACTT	AGCCAAATGA	GTCTCTCTGG	CAATTCTGCC	TTGTTTCAAG	ACTGGCTCAG	1800
TTAAATTTCT	TTTATCTTCC	GCCGGGCGCG	CTGGCTCACG	GCTGTAATCC	CAGCACTTTG	1860
G3AAGCTGAG	GCAGGAAGAT	CACCTGAGGT	CG3AGTTG	AGACCAACCT	GGCCAGCATG	1920
GTGAAACCOCT	GTGTCTACTA	AAAATCCAAA	CATTAGCCAG	GCGTGGTGCG	AGGCGCCTGT	1980
AATCCCACT	ACTTGGGAAG	CTGAGGTGAG	AGAAATCGCTT	GAACCCAGGA	GAGGGAGGTT	2040
GCACTGAGGC	GAGATTGTGC	CATTGCACTC	CAGCCTGGGC	AACAGAGCGA	GACTCCACCT	2100
CAAAAAA	AAGGATTTCT	CTATCTTCA	AAAACTCTAA	TGTTTAAACA	GSTCTTACAG	2160
TTGATCTAAT	TCAATCTCAT	TTTTTACAA	T3AGAAACA	G3ACAGT3A	CGGTGGATCA	2220
A3TGACACCA	GTAAGACTGA	GCTAAATTA	AAACGAGATC	TCACTCGAGT	CTGAGGTTAT	2280
TOCCACTGTC	CAACCTTACT	TTAAAGTAG	TTCAAATTTT	ACTTTTACTT	TTCCATAAAT	2340
TOGGAAGGGA	TTTTCCCTAG	GAGTCCAAAT	GTTGAAACCT	G3AAGGSTAT	AGTCTCTGTG	2400
TCTTTGAGAT	GAGGGGAGCC	CTGTCCATAT	TCAAGTTATC	AATTGACTTT	GTTGTTTTTG	2460
AGAAACGAT	CTGATTGGG	TAACTTTAA	ACATCTGTTT	GATTAGTCCT	ATAAAATATG	2520
CATATATAGA	AGACAGAAAG	AGCAACAACA	AATTTGAAAG	ATGCTTGTTA	AGTAAATCT	2580
GTATCGTAGG	TGTCCATTCC	TGCCAGTACC	TTTATAGTAT	GTAAGTTTAC	GTGCTGTAAT	2640
AGTATTAAATA	GTATCTAGAA	AATACTACAC	ATGCACAGCA	GTGCTAACTT	TGCCTTG3GA	2700
GTTGGAAAT	ACTTCAGAGA	AGCCAAACAG	CAGATTTTTT	TCTCTTCCCT	TCCCTTCTA	2760
ATTTTCCCTT	TCCCTTCTAC	CCCTTCTCT	TCTCTCCCA	AGTAACACTG	TGCACATATG	2820
TCAAACGAAA	ACTTATAATC	AAGTAACTGT	TTCTGCAAAA	ATAAGTTCGT	TTTCTGTCTA	2880
TGGCTCAAGG	CCTCAGCAGA	TCCAGGCTCT	GT3AGCG3GC	TGCTCTTCGT	CGTGTGCCAA	2940
ACACTGACCA	CTGCCCTGCG	TCTGCCATCT	TAG3CTTAAT	GAOCTGGGCTG	TTACTAA3CA	3000
CTGTCCCTCT	TGCCCCATGC	AGCTGTCTCC	TTCTAGTCTT	CTCCCTCTTC	TCAACCGCAT	3060
CCTAGCCCTT	CAG3CCATTT	CACCTCCATT	TTCCCTCACT	TCCCGCCGCT	CCTCCGCACT	3120
TCCTCCCTAC	TGTTGTTTCC	GCCCCACTAG	AGCCCTCAG	AGAAAGTTTC	CATCCTCGCA	3180
CCCTTCCCTG	TGTCCAGGCC	CGTCACATT	TCACAG3CGC	CCATCCCTCC	AGCCCCACCC	3240
CAAG3CCAAAT	GTACTTGGCG	GTAT333GAC	CTTCTCTGCT	AGCGAAACGCG	AG3AGT3AA	3300
GACCTG3GC	GCG333T3CT	CG3ACTTCC3	G33T33AG3T	G33AAG3CG3	CGGCACTCCC	3360
AGGAGCCCT	GCACAGTCA	CGTGAC3CT	CTCCACCCAC	CA3CCCGCC	AACTTCCCA	3420
CCGTAGCCTC	CCAGAGCCAG	GCCCCACGGA	AA333AGCTT	TTCCCGGTT	TTCTCCCGCT	3480
CTTTCCCTCT	CACTTGGAAT	ACTCGTGAAT	CAAAAATCTC	TCCCTGCCAC	CCTGTGT3TG	3540
TTTGAACCCG	GAAAAATCT	GAAACTG3TC	AA3AAAGAACT	AA3AAAGACT	TGCCAAAG3CA	3600
AG3CCG3TGT	GTGTCCCGAG	AGCTTAGAAT	CTCAGCAAG	GAACACAAAA	TAGCACATCT	3660
ADG3CCTCTT	TTGAGTAAA	ATTTACTT33	TTT3TTT3CA	G3AAGG3TTT	AAAACCTGCT	3720
TT3CAGATGC	TCTGTTT3CA	G3AAG3CCTT	AATCAAGTGT	TCCCTG3GC	CACAGCAAG	3780
GTTTTTATAT	CCAGAGCCTC	AGTTACTG3C	CCCTCTTCT	CTTGTG3GCA	AGCAACGTT	3840
CAGAACTCAG	CCTTCTTAGA	AAATTCTTAC	CCCG3GCTG	TCAATAAGTT	AAGTCTAATT	3900
G3CAACAGCT	ATCAAAAAGT	GTTGCATAC	ACACATG3CT	CACATAATTG	TAGCTTTG3C	3960
TCATAGG3TG	TTTTAATGCG	GAGGCTTTGA	CCTCAATCT	CAAGATATA	CATTCCAAAG	4020
TTAGC3CCAG	TTAGTGGATG	T33AA3AAAA	AAAAAAGCAA	ATTACCTCAT	AACACAAAG	4080
TCATAAACAC	ACATCCATAA	GTTCCAG3TA	CAAAATCTTA	CATCTTAGAG	AACTATATTT	4140
AACATTTACA	TACATTACTA	AG3TTTTTTT	TTTCTTTTGT	CTTGATTAAA	T3TTAGTTAT	4200
CATTAACTCT	T3GAATTATT	CTGT3T3TGT	ATATTTATTT	GTT3TTT3TG	AA3AAGCC3G	4260
TT3TTTTAAA	TAACTTCTTA	GAAAAATAG	GCTCAATCTG	TTTAATCTGA	GTTGCTAATA	4320
TT3TGAATA	TAGGCCACAT	AATACTAGCC	TAGATACTTA	T33CGAAGTA	AG3AGTCTCA	4380
AACACTGTCT	CAGAACATAA	GCAATCTGTG	TT3AATTTTT	ACCTCTGTGT	GTAAAATGAA	4440
G33AAAG3GA	ATGAAGTTTT	AGTTT3CCTT	AATTTTATCT	TTTATTTGTT	CAGACTCTTC	4500
AGCAGTATAA	AGTTTTCTAT	AAGTCAATA	TATTCACCTT	AAAGTCACTG	TGCTTTTATTC	4560
TGATACCAGT	TCTTCTCTAA	TTT333333	CAG3TGAGAT	AGTTTATATG	AAATAAAAG	4620
ATTAAAAAT	CTTACATTTT	TAGTCTCTCT	CCTTGGTAAA	AT3TAGAGTT	GTCCACTGTG	4680
TTTATCTCT	CCTCCTTATT	ATCAT3GTTG	CT3TTATTAT	TTTTAATG3T	TCATTAAACC	4740
CAAG3TCTG	G3AAATACTC	AT3GAATTTCA	TCTCAGAGCC	TT3CAGACTGT	ATGATATTTA	4800
AACAG3TGT	TGTCCATCTG	ATTCTTAAAA	TATTTCCAG	AAAAATGATT	CCACCTAATG	4860
CATAAATGCT	TTCATCA3AT	TAAG3AA3A	CCATG3ACAT	TTTATTTTAT	TTTATTTTAT	4920
AAATATTAAC	TTCCATTGCA	TAAGCTAAAT	G33TAGGAAT	AA3TGAGATG	ATATTGTTAT	4980
CTAGAGCTTT	AAAATATTCA	AAGGGCTGTC	ATCATTATCT	CATTTAATCT	TTGAAAACAA	5040
CTCTATGAAG	TACAAAG3AC	ACTGAGACAT	TTGTTGCTCT	ATATCAAGGA	AAAAAGTGTT	5100
TGTCCCAAAA	CTTCAAAATG	TGTAAATTAC	ACATTCTGCA	TCITTTACAG	TGGAGAAAAT	5160
TCACTG3CAA	TGGAATATTT	AAAATTA3AG	CTTGCTTAGT	GTGCTGCTTC	TGATCACTAC	5220
TTGATCCCA	TTCTGTCTTT	CATGTTAATT	G3CCCAATTG	GACTCTACAG	TTGGAAGGTG	5280

AAAACCTTACT	ATTTTCAACTT	GASTCACSTA	TGTATTETTA	TCATATACTT	CTTAAAGSTA	5340
CTATTTTTTTT	TCTTCTGATA	GTACACACAC	CAAGCACTTC	CAGCCACCCCT	GCCACAGACT	5400
TCCTTTGTAA	TCACTGTTGA	AGBACATBAT	GTTTTTATGA	CTTCCCGAAA	TGAAAACCCCT	5460
ATCTTGTTTT	TAAAACAAC	AAACCAACAA	AAAGTASTBT	TTATGTAAAG	ATTTTSTTCC	5520
CTGACTCTAG	GAACCCCTCT	GTTTTTATAT	CAACTCTGTA	CTGCAAAAC	ACAAAACAA	5580
AATGCCACCT	TGCTAATTCC	CTTCTAGCA	AAGTAATACA	GTTTAGCACA	TSTTCAAGAA	5640
AAAAATGGCT	AAGAAATTTT	GTTTCCACTA	ATTATTTTCA	ABCTGTGAT	ATTTACACTC	5700
TGCTCTTCAA	ACGTTACATT	TTATAAGACT	ATTTTTTAA	ATGTTGAACA	TAAGCCCTAA	5760
ATATATGTAT	CCTTAAATTT	TATTTCAAA	ATTTTAGTTC	AGTCTTTGCT	ATCATTCAG	5820
GAATAGAAAG	TTTTAACACT	GGAAACTGCA	AGTAAATATT	TGCCCTCTTA	CCTGAATTTT	5880
GGTAGCCCTC	TCCCCAAGST	TACTTTCTGT	TGCAGAAAT	GTAAAAATTA	TTACATAAAA	5940
TTCTAATGAT	GSTATCCGTC	TGCTTGTGAT	CTGATACAG	AGATAAAGAA	GTTTTATGAA	6000
AATGGAATCC	TGTTCCACTG	AAAASTAAAT	CTTAATGBC	TGATCAACT	ATCCTTTGAC	6060
ACCATATTGA	GCTTGGGAG	AAGGGGAAGT	CCTGAATGAG	GTTATAAAAT	AAAAGAAAT	6120
ATTTGCAAAA	TSTTCCCTTT	TTTAAATG	TACATTTTAG	AAATATTTTA	AGTGTGTAA	6180
CATTGTAGAA	TAGTCCCTAA	TAGGACTGAT	TATTCGCTAT	TGTAAATATA	GAAGAGTGT	6240
TGTGCTGAA	TGTGACCAAG	AAGTCTGAAA	ATGAAGAGAG	ACAATATGACA	AAAGAGATG	6300
CTTCTAATG	ACTAAGGAG	TGCTTTCTTA	AAGTCAAGAA	GAGATACTCA	GAAGAGGTA	6360
CAAGTTTTT	AAGGCACACA	GCCCCAACTT	TTAGGGAAGA	AAAATTTTCA	TGAAATAGT	6420
GATATTACAT	TAAAGAAAT	ACTCGTATCC	TCTGCTCACT	TATTTGCACT	TCCATTGCCC	6480
TAGGAAGAG	CCTGTTTTGA	GCGGGGCCCA	AGGASTGCCG	ACAGCACTCT	CCTCCCTCCA	6540
CCTTCTTCT	CATTCTCTCC	CCAGCTTCT	GAGCTCTTTC	CTCCCTGCTC	GACTGCTTGG	6600
ACAGTCAGCA	AGGAATTTCT	TCCCAGTCTCA	TTTTTCTCTC	CTGCTCTCCA	ACTCTGCTCTG	6660
CTAAAGCGGC	TGCCACCTGC	TGCTCTCTAC	ACAGCTCTCG	GAAGAGGAAA	GGAACTCTAG	6720
ACCTTCCACA	TGCTCTCTCC	TGCAACAAA	CTATTTGTCT	CAGGTAAGAA	ATATCTTCT	6780
TCTTTATTTT	GAAAGTCAGC	CATGSCAATT	AGAGGTAAAT	AAGGTAGAAA	GCAATTGAGA	6840
GGAAATATAAA	CCATCTAGCA	TCACTAGGAT	GAGCACTGAG	TATCAACATA	AGAAATATAA	6900
GCAAGTCAAG	AGTAGAATTT	TTTTCTTTTA	TCAATATATG	GAGATATCA	CTTTAGAGGA	6960
GAGGTTCTCA	AACTTTTTTC	TCTCATSTTC	CCTTTACATT	AAGCACATCA	CATGTTAGCA	7020
TAAATAACAT	TTTTAATTAA	AAATAACTAT	GTACTTTTTC	AAACACAAA	AAAGGTATAA	7080
AGAGTGACAC	TTTTTTATTT	TTACAASTGT	TTTTATGCT	TTAATAGAA	CCATATAGAT	7140
CTGCTGGAAT	CTCATCTGCT	TTGCATTGAG	ACTACTGCAA	TATTGACAG	AATGAGGCT	7200
CTGCTAAACT	CTGTTGTACA	CTCATGAGAG	AATGCTTCAA	AAAGACAAAT	TACGCTTTAG	7260
AATTATTAGA	AATAGCTTTC	ACTTTAGGAA	CTCCTGAGAA	ATTGCTGCT	TAGAGTGTAA	7320
AGATAAATAA	GCTTCTCTTT	AAACGGAAAT	TCAAGACAGA	ATCAAGTTACA	TTAAAGGCAA	7380
ACAAAATAAT	TGCCCATGCT	TAGTCTCTT	GTGAATCTG	CCACACCTTT	GGACTGGBCT	7440
ACAATTGAT	AATATAGCAT	TCCCCGAGAT	AATTTTCTCT	CACAATTAG	GAAGGGGCTG	7500
AATAAATATC	TCTGTTTGAA	GTTGAATAAC	AAAAATTAG	AGCCCTTAAA	TTTTAGGGBCT	7560
CCTGAAATTC	GTCTTTTTC	CTATATTGAG	CTACTTTAG	TTCTATTAAA	TCTTTTTC	7620
GCTCAGSTCT	ACTAGCTCAT	GCTTAGATC	CTAGGCTGCT	CTGAGGCTAG	GAATTTGAGA	7680
CCAGCCAGGG	CAACACAGTC	TCTACAAAA	AATAAAAAAT	TACCTGGGTC	TGTTGGTGCA	7740
TGCTGTAGAA	ACTACTCAGG	ATGCTGAGGA	CTGCTTGAAG	CCAGGATAG	CAATCTCTG	7800
GTGAGTTCAG	CACTAAGACA	GAGCGAGACT	TTCTCAAAA	AAACACAAA	AAACACAAA	7860
AACTTCTCTC	AAAATAACTT	TTTATCTGCA	ATSTTTTCT	ATGCTCTGTC	AGATTAAAT	7920
TACTCTTTTA	CCTGATTTTC	AAAGCTCTCT	ATAATTAAT	CGGCTTTTAC	CTTGTGTTCA	7980
CTGCAAAATA	GTAGGACTGT	TCCACTACAA	TCCAAAAATC	ACAGTTTGG	TGCASTGGBCT	8040
CACTCTCTGA	ATCCCAACAC	TTTGGAGGG	CAAGGCAAGT	GGATTGCTTC	AGCTCAGGAG	8100
TTCAAGATCA	GCTTGGGCAA	CATGBCAAAA	ACCTTCTCTC	TCAAAACAT	ACAAAATTA	8160
GCCAGATGTC	GTAGTATGTC	CCTGTAATCT	CAACTACTCA	AAAGGCTAAG	GAAGAGGAT	8220
CACTTGAAG	CAGGAGGTC	AGCTTACAGT	GAGCTATGTT	TACTGTGTCA	CTGCACTCCA	8280
GCTTGGTGA	TAGAGCAAGA	CCATGTCTCA	AAAAAATAA	AAAGAAAGA	AAAGAAAAA	8340
ACATGGGCT	ATTCASTTCA	CCCCCAACAC	AACATTSTTT	TGATTATCAC	ATAAATGCTG	8400
GTCCATTGCT	TTCTCTATCT	ATTCAAATCT	TTAAGCATTC	TTGAGATTTC	AACTCAATTC	8460
TCTTTTCAA	ACTAGGCTAT	TTAAACTACA	TCASTTCTAT	TTGATTTTC	TTGTTTTGAG	8520
TCTACAGCT	CAAAAACAAA	AACTTAAAAA	CTTATTTTTC	AASTTTCTG	CTACTCTCAC	8580
TTCTTCAACT	CTCACATACA	CACATTCTATA	ATAACATGCT	AGAAATGTTCA	AGGATAAAT	8640
GATTTATAGA	ACTGAAAAAT	TAGTTTTGA	TCTTSTTCT	GTCAAGATGA	CTACTACTCT	8700
GATCTGAGST	AATTAATTAT	GTAGCATGCT	CCCTCATTTTC	ATCCCATAG	TATTCAGAG	8760
GATTGGAAT	CCACAGCAAG	GATAAATATA	ATCATATTC	CTTTTCAAGT	TCAAGGCTATT	8820
TTAACITTTA	ATCTAGTAGT	ATGTTTCTT	TTGTTGTTCT	TGTTTGAAT	GGAGGCTGCT	8880
TGTCTACCTC	AGGCTGAGT	GCACTGCTAC	GAAGCTGCT	CACTGCAAG	TCTGCTCAT	8940
GGTTTCAATC	AGTTATTCTG	CCTCAGTCTC	CCAACTAGCT	GGACTACAA	GCTCATGCTC	9000
ACCATGCTCT	GCTAATTTT	GTATTTTGA	TAGAAACAG	GCTTCACCAT	GTTGCTCAGG	9060
CTGCTCTCGA	ACTCCTGAG	TCAAGTGATC	CAGCTGCTCT	GGCTCTCCAA	AGTCTGCTGA	9120
TTACAGGCTAT	AAGCCACCT	GCCCCGCTA	ATAGTATTT	TTTAAACTCT	TAGTGTCTTA	9180
ACAAATGCTG	TGTATAATA	AATATGCTAT	AAATATTTAC	TGTCTTGA	TTATGAAGAA	9240
GTGTTTACTA	GGCCGTTTT	CACATATCAA	TGTTCTCTCT	CTTACAGTT	TAATTAAGAT	9300

CTAGAATTG	AGGTTTGTAG	AGCTGGAACA	GACCTTAAAG	ATTGACTAGC	CAACTTCCTT	9350
GTCCAAATGA	GGGAACTGAG	ACCCTTAAAA	TTAAGTGAAT	TGCCCAAGAC	AAAACTG3AA	9420
CTCATGTGTC	CTAATTTCCA	TCATGAAATT	CTACCATTCA	CTA3CCTCTG	GCTAGTTTCT	9420
AAAGTATTTC	ATAACTAAAT	TTTTATTTCT	GTTTTAAAGA	ACAAATTGTC	ACTGCTTACT	9540
CCTG3GAG33	TCTTTCTGAG	GTGGTTTATA	ACTCTTAAAA	AAAAAAAAT	CAGTAGTCTG	9600
AGAAATTTAG	ACGAAATAGT	CAAGCTATTT	TTATCCAATG	GATCTATAAT	TTTCATAGAT	9660
TAGACTTAAA	TCAAAGAAAC	ACGATGAGA	AAGGAAGAGG	AAATTT3AG3	AGAGGAG3AA	9720
TG333ATGAG	AACACACTAC	TTGTAATCAG	TCATAGATGT	ACT3AGAACT	AACAAGAGAA	9780
ATTGTAG3AA	AATAAGAAAT	AAGAATTCAA	AATCAACACA	TGAAATAAAA	AGAACTACT	9840
A3G3AAAAAT	G3AGAAGACA	TTAGAAAAAT	TATTTCTATTT	TTAAAAATCT	GTTTTCA3GC	9900
TTCCCTCCTG	TTCTTCCTCC	TTCTCATTGG	TTTTCA3GTG	GAG33AAAT	TTAAGATGGA	9960
AAAAATATAT	ATATTCTACA	CATCCCTTTT	TAGCTGTGTT	TCAT33CAAC	AA33TTTATC	10020
ATAG3AAACT	TTTATTGATA	CAATATTTAT	TGAGTTCTTA	CTGTGT33TA	AGCTCTTTCC	10080
AG3TGTG3AA	AATTCAGGGG	AAAAAAGACA	ACTCATTGTC	TTAAACTCA	GATGAAAGCT	10140
GAACAGACCT	ATTTTAAATC	AAAGTAATCT	CAATTTAGGG	TAGTAAGAGC	TATTTAAGAA	10200
GCATGAAACG	GTGTGAAGCA	G3TAGSACTC	TGAG3AGAGA	ATAGTTAGCT	AGGAATGAAA	10260
GAG3AGAGAA	GTTTTCTAG	AGGAATATT	AAAGCTGGGA	GTTAC3GGAT	GAAAGATGAG	10320
GCA333TTTT	CAG3CAAAAA	AAAAAAAAG	GCAG3G3AAG	GGGAAATTTCT	GGCCTGGCAG	10380
AGAAATAAAC	TGTGGCAACA	ATG3AG3AGA	GTCTGGAAGC	AAGAAAACCA	AGTAGAAGAG	10440
TATTAATAAT	GAAGATGCCA	G333TAATGA	G33CTTGATT	TAAACAGTG	CTGTTGGAGA	10500
T33AG3AGAG	ATAACAAATT	CTGGAGACAT	TTCTGAGTTA	GAACTACAG	TATTTATCAG	10560
ACAAG33AAA	GATTAGACAA	AGGAGTTAA3	AATGACTCCC	AGGTTTCA3T	TTGGGGG3AG	10620
TAACTAG3AC	ATGTTTTGAA	AAGTAATGTA	TTGGATCTCT	TACCATTG3A	ACTATGTATG	10680
TGGAGCCAAA	TTAAATTTTG	TACATGTATA	TAACCTCTCC	CCCACCAACA	GTAACACTT	10740
CCCTAACTCT	CTACTTTGTA	GCCAGACTTC	CTAAAG3AAT	AGTTT3TAGT	CAGTGTCTTT	10800
ACTTTT3333	TCCCATCTCT	TCTAGATAT	TTGTCCACCT	ACCATCTGCT	GCCTCCACTT	10860
TACCCAAACT	GTTCTACGGT	TGCCCCAAAC	TTCTTAATTG	CCAAATTCAA	TGACCAAGTT	10920
TAACTTTATA	TGTATATTAG	GAGCTCTACA	GTTTGATTTT	GAGCAGCCCC	TCCT3AAACC	10980
CTTTCTCTTT	CGATTTCTGT	GACACATCTC	AGATTTACAA	AACTCAACTA	ATTATTTTAC	11040
ACTT3AGCTG	TATTTTCGTT	CTTCTTTCTT	GAT3AATGAG	GTAACCACTC	AGCAATTTGC	11100
CCAA333AAA	AACTACGAAG	TCATCTCA3	TTCTCTCTTC	TTCT3TTTGA	CCCACCAACG	11160
ATCAGCTGAG	AAATCCCGCT	GTTTAGTATC	TCTTGAATTC	ATTACCTTAA	TTTATAGCCT	11220
CATCAACTCT	TAATTGTTAA	AATTACTTCA	GTAGTTGTTG	TCTGACCTCT	GTCCAACTCT	11280
GTTCAATCAG	GTCCATTCTT	TTGTTCTTGG	TG3T33GTG3T	GCTGTT3ACA	GAGTTTCTGCT	11340
TTTCT3333	AGGCT3AAGT	GCAGT3GAG3	ACTTCACTG3	AACTAG3G3C	TCCTGGG3TT	11400
AAGCA3TTCA	CCCTCCCGAG	TAGCTG33AC	TACAG3TAGT	TGCCACACACA	CCGAGCTAAT	11460
TTTCT3TTTT	CAGTAGAGAC	AGGGTTTTCAC	CAT3TTGGTC	AG3CT33TCT	CAAACTCCTG	11520
ACCTCA333A	ATCC3CCAC	CTCAGCCTCC	CAAAGTGCTG	GAATTAAG3G	CATGAGCCAC	11580
TGCACAC33A	CCAGATCCAT	TGTTTAT3TT	GTTTCTAGAG	TGATTTT3TA	AAACACAAAT	11640
TTGACCATAT	CTTTCTCCAA	TTTAA3TCA3	TATTTT3TTT	TTCA333AAA	AAACGTTCAA	11700
ACTCTTTAGT	CTGCTTACAC	AAGCCTTTTG	TATCTGACT	CTTCTTTCCA	AGTTTCTATC	11760
AAAGTATACT	G3AAGTTACA	TTTTAT3T3A	ATTGAATTAG	GCAAG333AT	AA3AATTATA	11820
GTTTATAT33	GCAAAATGGA	AATAATGTTA	ACTCTTCCAA	ATA3TTTATC	TAGAATGACA	11880
TAAATTTCAA	GCTGTACAGT	CAAT3AGTT	ATAAACTGTT	AACACTATTG	CCACATGCCA	11940
TGTCTCTCTA	TACTTGGTAG	AATTAT3T33	TTCTAGTCA	TTATTATGTA	AATTAGACTT	12000
TAAATAACT3	AGAAGTTCTT	CAGACATACA	TTCTTATATT	GTGCTTTTAA	AACTAATTTT	12060
TAAATAATTT	TATATATGAT	AATGTTATCC	AA3T3CTAAG	GGAT3TATT3	TTACT3CTGT	12120
GCAAA3333A	AAAAAAAATA	AACTCCAAAT	AAATATGTTG	AAACCAAGTT	TATAT3CAAG	12180
AAACCAATAT	TAAAGAGGCC	AAAGTAACAC	CATAATAG3C	TGT3T33AGA	CG3CA333TA	12240
CAAAACACTA	GTAATAAT3C	TGAGAA33TT	GAGAA333AA	AGAA333AAC	AATAT33TTT	12300
GTTT3TT3TA	G3TTTATGTA	CTCCAA3AAT	ATCTCTCTC	AAAC3TTTAC	GTTTTTTTCA	12360
AAGAA333TT	AACTTTGGCT	GG3333AGTG	GCTCTTGGCT	GTAT3333AG	CTTTT333AG	12420
GCCAA33333	GCAGATCACC	TGAGGTCA33	AGTTTGAGAC	CAGCTTGACC	AAAAAT33AG	12480
AAACCC3333	CCCTCACTAC	TAAAG3AATA	CAAAATTAGG	CCG33333AG	TG3CTTACCC	12540
CTGTGATCC3	AGCACTTT33	GAG33333AG	CAG3AGAGTC	ACCT3AGT3C	AG3AGTT33A	12600
GACCA3333AT	GSAGAAACCC	GTCTCTATTA	AAATACAAA	ATTAG3333G	CTTGGT333T	12660
CATGACT3TA	ATCCAGCTA	CTCAG3AG3C	TAGG3AGAG3	AATCACTTGA	ACCCAG33AG	12720
TGGAGTT333	AGT3AGCC3A	GATCCT33CA	TT3CACTCCA	GCCT3333AA	CAAGAG333A	12780
ACTCT3TAT3	CAAAAAACAA	AAGAAAG3AA	AG3TAACTT	T3AACTATGT	GAGATCTTTA	12840
GAAAT3CATT	CTTTCTGTAA	AATGT3ACTA	CATTT3CCTT	ATTTAT33TA	AAAT3TT33A	12900
GGCTCTCAAC	AACCCATATT	TTCTG33TCT	CCCCGCT3CC	TABCTTTTGT	TCACATT3CT	12960
TCTTCTT33T	GGAA3CTCTT	CCTCT333CT	TGAAAAT3CC	T3TTTCTCTT	TCAAG3TAGC	13020
ACASTCATCA	CTTTCTGT33	TAACCTT3TC	CAG3ACCACT	AAACAGAAAG	AATGAATCTC	13080
TT3TAAATTC	AGCTCTTAAG	TCATTCAATTA	CATTATTTT3	TAACTCTTTA	TAGATCTTTC	13140
TCTCCCACTA	GACTCTGAGT	CAGT3AGAG3	TAG3AGCCAA	CTCTCAATTA	TGTGTGGTTT	13200
GGTCAGCTAC	TG3CCACATT	CCTGAT3CAT	AGTTAATGCT	CAAACTTTAA	CTG3TGAATC	13260
AGCTCAATA	TTGTCTTCTT	CTAAATCCAT	TCATCTATTG	ACTAACTATG	TACTCAAAAT	13320

AGTAAACACC	AGTAATTTAA	TCCAAATTCCT	GCCCATACTG	CTTGGTACAT	TTGAGGTGAA	13380
TTAGTTTGTAT	AAATATGTGT	GTATTACATA	ATATTAAAGT	ATGTACAGAA	GATCATGCTA	13440
ATCATAATTC	ACAACTGATA	ACTAATCAAA	CATAAATGCT	CTCAGGTAA	CAAATGTCTG	13500
CCTTCTCAGT	TAATGCAGTC	ATTAACAAAC	ACCTTCTGAT	GCTGATAATA	GGSCCTTGTT	13560
CAGCAATGAA	GCCATAAAGG	TGAATAAAGA	ACATGCCCTC	GTGGAGCTCA	CAGCCTAGTC	13620
ATTATTGTTT	TGATTTTTAA	TATTAATGTT	GGTTTGGGTT	TTGGTGAATA	ATGTTTAGAC	13680
TTATCTTAGT	GATCTTTTCA	TCCTTTGCTA	TATTATTTTT	CTCTAAGAGT	CTTCTTATC	13740
CCCTCCTTTA	AAAAACTAGG	TGATAATTCT	AAATTGTAAA	TTTAAATATT	ATAAATAGCT	13800
TATAAAATTT	AATATTTATA	ATATTTAAAT	GTTTGATAAA	TATTTAAAT	TTATAATATT	13860
TAAATGTTTT	TTTAAATTCA	TTTGACATC	ASTTTTTATT	TTATTTAAAT	GTGTTGGCCA	13920
G3CATG3TG3	CTGACACCTA	TAATCCGAGA	ACTTTGAGAG	GCCAAAGTCAG	GCAAAACCATT	13980
TGAGCTCAG3	ASTTTGAGAC	CACCCCTGGC	AAAGTGGTGA	AACCCCTGTCT	CTACCAAACA	14040
TATGAAAAC	TATCT333T3	TGGTGGCAGG	CATCTGTGTT	CCCAGATG3G	A3TCCCAG3C	14100
TAAGATG3GA	GAATGCGTTG	AACCCCAGTG	AGAGGGGTG3	G3T3GATGTT	G3AGTGAG3C	14160
GAGATCGTGT	CACCTGCACTC	CAACCTGSSST	GACAGAGTGA	GACTCCATCT	CAAAAAAAA	14220
AAATGTTATC	TAAATAAGAT	AAATTTAATA	ACTGTTG3CA	CTTAGAT3AG	CATAAGGAAC	14280
TAAACCTAGA	TAAAACTATC	AAATAAGGCC	TGGGTACAGT	GACTCATGCT	TGTAATCTCA	14340
AGCACTTT33	GAGG3CAAAA	TTATACAAA3	TTAGTTGTAT	AACACCAACT	AACAACATATT	14400
TTGGGGTTAG	CTTAATTCAG	ATTAATTTTT	TTTAAACTGA	GTTTTAAATT	CCTGCTTACT	14460
CTACCATACA	TGCTAG3CCT	CATATTAT3C	TAGAAAAATT	TTGAG3CACAG	ATTTATGAAT	14520
ACTCTCCT3C	ATACCATTTA	ATTTTTAAAC	AAATTTTAAT	G3AGTATATA	TGTGCTTTTT	14580
TACCAACACA	TTAAATAATA	AGATCTACTG	TGAGGACTAA	ATTTCTGTAA	TTTCAAAGTA	14640
GTAATGAGTT	TAAACCATGT	CTCAAGATCT	CTGCAATAAC	TGTAG3CACAA	CAGAAAATAG	14700
GTATTTCTAT	TAAATGACAGA	GTCACAAAGTA	CTACTAATAA	TACT3TG3TT	TGTTTTCTCTG	14760
AACATAATCAT	GGAG3GAAT3	CTAAATTTCA	GAG3TTG3TG	AAAATACATG	TGTATTTTTT	14820
TCCCATC3CA	AGTT3ACAGA	TTCTCTACAC	TGAGAACTCC	TATTC3CATA	CAAAATCTCTG	14880
GAAGCCT3CA	CACCT3TATG	GAAGAAGGGC	AGAAAG3AAA	AG3CAATG3A	AGGATTTAAA	14940
TTTTTTT3CA	ATCCTGTATC	CCTTGATTTT	ACAGCAAGAT	TGTATTTATG	TATTACTTGT	15000
GTTAAAAATA	TAGTATAATC	GAGACTCCAG	ATCAAAAATC	ACCG3AGCTC	AGGGAGAAAG	15060
AGGGCCAC3CA	AATGCCAGAG	CCCTTCAGCG	TTCTCCCACT	CT3CCTGTAT	CCTCAGATGG	15120
AAGCACTTTT	TTATCCTTTG	TTCACTTTTA	GCTTTT3GAC	AATGAAGTCA	CAAACTTCA	15180
GCCTCTCACC	CATAG3AACC	CACCTG3TTG	AAGAGAAG3GA	T3AA3CCAGT	CCTTCTTAAA	15240
GG3CACGATT	AGATG3GTTT	AT33CATCCT	CAG3TGAAAC	TATATTTATA	TTGACAAAT	15300
ATTTATATTT	CTCAAG3AAT	ACTAG3AATA	TGATTCAGTT	CAGTACTAG3	CCATTTATCT	15360
ACCTTTTATA	ATATTTGTTT	ATGAG3AAAT	GCTTTCTATC	TTCCAAATAT	CTGATGATTT	15420
GTAAGAG3AA	ACTTAAACAT	G33TATTCAT	AAG3TG3AAAC	TTCT33CAT	TATTGAATGT	15480
CAAGATT3TT	CATCAGTATA	CTAG3TGATT	AACGTGACCAC	T3AACTTGAA	G3TAGTATAA	15540
AGTAGT33TA	AAAG3TACAA	TCATT3TCTC	TTAACAGAT3	GCTCTTTGCT	TTCAATTAGGA	15600

ATAAAG	ATG	GCT	GCT	GAA	CCA	GTA	GAA	GAC	AAT	TGC	ATC	AAC	TTT	GTG	GCA	15651
Met	Ala	Ala	Glu	Pro	Val	Glu	Asp	Asn	Cys	Ile	Asn	Phe	Val	Ala		
-35										-30						

ATG	AAA	TTT	ATT	GAC	AAT	ACG	CTT	TAC	TTT	ATA	G	GTAAGGC	TAATGCCATA	15702		
Met	Lys	Phe	Ile	Asp	Asn	Thr	Leu	Tyr	Phe	Ile	Ala					
-20																

GAACAAATAC	CAG3TTCAGA	TAAATCTATT	CAATTAGAAA	AGAT3TT3TG	AGGTGAACCTA	15762
TTAAGTGACT	CTTTGTTGTA	CCAAATTTCA	CT3TAATATT	AAT33CTCTT	AAAAAATAG	15822
TGAGCCTCTA	GAAATTAACC	ACAACATGTC	CAAG3TCTCA	3CACCT33TC	ACACCACGTG	15882
TCCT3GCTACT	TTAATCAG3CA	GTAG3TCACT	CTCAG3TTG3	CAGTAAGTGC	ACATCATGAA	15942
AATCCCACTT	TTGATG33GA	AATCCCACTT	TTGATTTGAT	TTCCATG33GA	AAAATCCCA3	16002
TACAAAACCTG	GGTG3ATTCA	GGAAATACAA	TTTC3CAAG	CARATT33CA	AATTATGTAA	16062
GAGATTCTCT	AAATTTAGAG	TTCCGTTGAAT	TACAC3ATTT	TAT3TAAATA	TGTTTGACAA	16122
GTAAAAATTT	ATTCTTTTTT	TTTTTTTCT3	TT333TAGGT	T33AG3T3AG	T33CACAAATC	16182
TCTGCTCACT	CCAACTCTCA	CCTCT333GT	T3AAGCAATT	CTCT33CTC	AGCCTTTCTGA	16242
GTAGCTG33A	CTACAG3T3C	AT333333AT	GGCT333CTAA	TTTT3333TA	TTTTTACTAG	16302
AGACAG33TT	TTG3TATGTT	GTCAG33CTG	GTCTT3GACT	CCT3ATCTCA	GATGATCTCTC	16362
CTGGCTCGGG	CTCCTAAAGT	GCT333BATT	CAG33CATGAA	CCACCACACA	TG33CTTAAA	16422
ATTGATTCTT	ATGATTAATC	TCCTGTGAAC	AATTT33CTT	CATTTGAAAG	TTTGCTTCA	16482
TTTGAAAAC	TCATTTAAAA	GCTGAG3CAA	CAAG3TGAGA	CCCCATCTCT	ACAAAAAAT	16542
GCAAAATATC	CTGTTGACAC	CTCCTACCTT	CTG3TGAGGC	T3AAG3AG3A	GSATCACTTG	16602
AGCCTAG3GA	TTTGAGCCTG	CAGTGAGCTA	TGATC3GAGC	CCTACACTCC	AGCCTG3CATG	16662
ACAGTAGACC	CTGACACACA	CACACAAAAA	AAAACCTTCA	TAAAAAATTA	TTAGTTGACT	16722
TTTCTTAG3T	GACTTTCCGT	TTAAGCAATA	AATTTAAAA3	TAAATCTCT	AATTTTAGAA	16782
AATTTATTTT	TAGTTACATA	TTGAAATTTT	TAAACCTAG	GTTTAAGTTT	TATGTCTAAA	16842
TTACCTGAGA	ACACACTAA3	TCTGATAAG3	TTCACTTTAT	GGSCCTTTTG	GATGATTATA	16902
TAATATTCTG	ATGAAAGCCA	AGACAGACCC	TTAAACCAT	AAAATAGGAG	TTCCAGAAA3	16962
AGGAGTAGCA	AAAGTAAAA3	CTAGAAATGAG	ATTGAATTCT	GAGTCGAAAT	ACAAAATTTT	17022

ACATATTCTG TTTCTCTCTT TTTCCCCCTC TTAG CT GAA GAT GAT G GTAAAGT 17075  
Ala Glu Asp Asp Glu  
-10

AGAAATGAAT TTATTTTTCT TTGCAAACTA AGTATCTGCT TSAGACACAT CTATCTCACG 17135  
ATTGTCAGCT GAGGAAAAAA AAAAATGCTT CTCATGCTAC CAATCTGCTT TCAAAGAAAT 17195  
GTGGACTCAG TAGCACAGCT TTGGAAATGAA GATGATCATA AGAGATACAA AGAAGAACCT 17255  
CTAGCAAAAG ATGCTTCTCT ATGCTCTTAA AAATTCTCCA GCTCTTASAA TCTACAAAAT 17315  
AGACTTTGCT TGTTTTATTG GTCTTAAGAT TAGCATGAAG COATGGATTC TGTGTAGGGG 17375  
GGAGCGTTG ATAGGAAAAA GGGATTGAAG CATTAGAANT GTCCAAAAATC AGTAACACCT 17435  
CCTCTCAGAA ATGCTTTGGG AAGAAGCCTG GAAGGTTCCG GBTGGTGGT GGGGTGGGGG 17495  
AGAAAAATCT GGAAGTAGAG GAGATAGGAA TGGGTGGGGG AAGAAGACCA CATTACAGAG 17555  
CCAAAAAGCT AAAGAAACCA TGBCATTTAT GATGAATTCA GGTAAATTCA GAATGGAAGT 17615  
AGAGTAGGAG TAGGAGACTG GTGAGAGGAG CTAGAGTGAT AAATAGGGTG TAGAGCAAGA 17675  
CGTTCTCTCA CCCCAGATG TGAATTTGG ACTTTATCTT GGAGATAATA GGGTTAATTA 17735  
AGCACAAAT GTATTAGCTA GGTAAAGAT TAGTTTGTGT TAACAAAGAC ATCCAAAGAT 17795  
ACAGTAGCT AATAAGATAG AGAATTTCT TCTCAAAGAA ACTCTAAGTA GGCAGCTCAG 17855  
AAGTAGTAT GCTGGAAGCA ACTGATGAT ATTGGGACCC CCAACCTCTC TCAGTCTTGT 17915  
ACCCATCAT CCCTASTTGT TGATCTCACT CACATAGTTT AAAATCATCA TACTTCCTGG 17975  
GTTCCATATC CAGTTATCAA GAAAGGGTCA AGAGAAAGTCA GGTTCATTCC TTTCAAAGAC 18035  
TCTAATTGCA AGTTAAACAC ATCAATCCCT CTCATATTTT ATTGACTAGA ATTTAATCAC 18095  
ATGGCCAGCA CAAGTGCAG GAAATCTGGA AATATAATC TTTATTCCAG GTAGCCATAT 18155  
GACTCTTTAA AATTCAGAAA TAATATATTT TTAATAATAT ATTCTGGCTT TGGTATAAAG 18215  
AATTGATGGT GTGGGGTGAG GAGGCCCCAA TTAAGGGTTG AGAGCCTATT ATTTTAGTTA 18275  
TTACAAGAAA TGATGGTGTG ATGAATTAAG GTAGACATAG GGGAGTGCTG ATGAGGAGCT 18335  
GTGAATGGAT TTTAGAAACA CTTGAGAGAA TCAATAGGAC ATGATTTAGG GTTGGATTTG 18395  
GAAAGGAGAA GAAAGTAGAA AAGATGATGC CTCATATTTT CACTTAGGCA ATTTGTACCA 19455  
TTCAATGAAA TAGGGAACAC AGGAGGAGAA GCAGGTTTTG GTGTATACAA AGAGGAGGAT 18515  
GGATGAGGCA TTTGGTTTTG GATCTGAGAT GTCTGTGAAA CGTCTAGTG GAGATGTCCA 18575  
CAAATCTCTC TACATGTGCT TCTGAGTTCA GGACACAGAT TTGGGCTGGA GATAGAGATA 18635  
TTGTAGGCTT ATACATAGAA ATGGCATTTG AATCTATAGA GATAAAAAGA CACATCAAG 18695  
GAAATGTGTA AAGTAGAGA GAAAGGCA AACTCTTGCT TGGGGGGAAT ACCTACATTT 18755  
AAAGGATGCA GTAGAAAGAA GCTAATAAAC AGACAGAGG AGACTAACCA AAAGGGAGAA 18815  
AGAAAAACCA AGAGAATTCC ACCGACTCCC AGGAGAGCAT TTCAAGATTG AGGGGATAGG 18875  
TGTGTGTGTG AATTTTGCA CCTTGAGAAAT CAAGGGCCAG AACACAGCTT TTAGATTTAG 18935  
CAACAAGGAG TTTGGTGAT TCAATGAAA CAATTTGATG GTGAAATGGA GGCAGAGGCA 18995  
GATTGCAATG AGTGAACAG TGAATGAAA GTGAAAGAA GATACAGATA ATTCTGTCTA 19055  
AAAGCTTGGC TGTAAAGAG AGGAGAGAAA CAAGACTAGT TSCAAAGTGA GATTGGGTTG 19115  
ATGGAGCAGT TTTAAATCTC AAAATAAAGA GCTTTGTGCT TTTTGTATTA TGAAAAATAAT 19175  
GTGTTAATTG TAATAATTG AGCAATGAA AAAAGATAAT AATATGAAAG ATAAAAATAT 19235  
AAAAACCACT CAGAAATAAT GATAGCTACC ATTTTGATAC AATATTTCTA CACTCCTTTT 19295  
TATGATATA TAGAAGACAA GAAATGCTTA TATTTTATTT AAAAGGGATT GTACTATACC 19355  
TAAGCTGCTT TTTCTAGTTA GTGATATATA TGAACATCTC TCCATGGCAA CGAGTAATTG 19415  
CAATTTATTT AAGTTTATGA TATTTTCAAA TAAAGGAGTA TCTTTGGCTT TTTTATTTAA 19475  
TCAATTTCTA ATTTGTTGAT GTTTGTTTCC AGTTTGTGTT TGTATTAAAC AATGTTCCCA 19535  
TAAGCATTTT TGTACACCAA TGTTCACACA TTTGTCTGAT TTTTCTTTCA GATATAAAAC 19595  
CAGGAGGTAG AATTGCTGG TTGATAGAA AGAAAGGATG ATTGCCAAAT TAAAGCTTCA 19655  
GTAGAGGGTA CATGCCGAGC ACAAATGAAA TCAGCCCTAG ATACCAGAAA TGGCACTTTT 19715  
TCATTTTCCC TTGGAACAAA AGGAGAGAG GCAATAAATG TGCTGCCAGA GTTAAATTTG 19775  
TACGTGGAGT AGCAGGAAAT CATTTCTGTA AATGAGAAAT AGAGATGATG TTGTAGAGGT 19835  
CCTGAAGAGA GCAAGAAAAA TTTGAAATTT CBTCTATCAG CTATGGAAGA GAGTGTCTAA 19895  
CTGGAAAAA AAAGAAATG TGAATTTG TATGCTTGA ATGGCACCGA TTTGAACGCT 19955  
TGTGGCATTT TTAACAGCA GCACTCAGCA GCAAGTTTGG GATTTTGTGA GCAGAAAGAC 20015  
AAATAAGTTA GGGATTTAAT ATCTTCTCCA AATGTTAGAC AAAATGAACT CTGAGATCCA 20075  
GCTGCACAGG GAAGGAAAGG AAGACGAGAA GAGGTGATAT AGGAAATACA AGAGTCAGGA 20135  
GACTGGAAGA TGTGTGATA TTTAAGAAA CATAGAGTTT GASTAAAAAT GTAAGAAAAA 20195  
TAGAAGGTTA AGAGACCGT CAGAAAGTAG GCTATTTGAA GTTAACACTT CAGAGGCAGA 20255  
GTAGTTCTGA ATGTAACAA GAAATTTAGT GTGCTTTTGA GAGTAGTTTA AAAAACAATA 20315  
GGCAACTTTA TTGTAGCTAC TTCTGGAACA GAAGATTGTC ATTAATAGTT TTAGAAAACT 20375  
AAAAATATA GCATACCTAT TTGTCAATTA ACAAAGAAAC TATGATTTT TAAATGAGAT 20435  
TTAATSTTTA TTGTAG AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA 20486

Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu

-5

1

5

TCT AAA TTA TCA GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT 20534  
Ser Lys Leu Ser Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe Ile

10

15

20

GAC CAA GGA AAT CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT 20582  
Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp Cys

25	30	35	
AGA G	GT ATTTTTTTT	ATTGCGCAAA	ATAGAAATGA CTAGCTACTT CTTCCCATTC 20638
Arg Asp			
40			
TSTTTTACTG	CTTACATTGT	TCCGTGCTAG	TCCCAATCCT CAGATGAAAA GTCACA33AG 20698
TSACAATAAT	TTCACTTACA	GSAAACTTTA	TAAGGCATCC ACGTTTTTTA GTTG333TAA 20758
AAAATTGGAT	ACAATAAGAC	ATTGCTAG33	GTCATGCTCT TCTGAG33CTG CCTTTGAATC 20818
ACCAATCCCT	TTATTGTGAT	TGCATTAACT	GTTTAAAACC TCTATAGTTG GATGCTTAAT 20878
CCCTGCTTGT	TACAGCTGAA	AATGCTGATA	GTTTACCAG3 TGTG3T33CA TCTATCTGTA 20938
ATCCTAGCTA	CTTG33AG33	TCLAGCAG3A	GGATTGCTTG AGGCCAG3AC TTT3AG3CTG 20998
TAGTACACTG	TGATCGTACC	TGTGAATA33	CACTGCACTC CAG3CT33GT GATATACAGA 21058
CCTTGTCTCT	AAAATTAATA	AAAAAAAAAA	AAAAAACCTT A3GAAA33AA ATT3ATCAAG 21118
TCTACTGTGC	CTTCCAAAAA	ATGAATTCCA	AATATCAAA3 TTA33CTGAG TTGAAGCAGT 21178
GAATGTGCAT	TCTTTAAAAA	TACTGAATAC	TTACCTTAAC ATATATTTTA AATATTTTAT 21238
TTAGCATTTA	AAAGTTAAAA	ACAATCTTTT	AGAATTCATA TCTTTAAAAA ACTCAAAAAA 21298
GTTGCAGCGT	GTGTGTTGTA	ATACACATTA	AACTGTGG33 TTGTTTGTTC GTTTGASATG 21358
CAGTTTCACT	CTGTCAACCA	GSCTGAAGTG	CA3TGCAGTG CAGTGGT3TG ATCTCG3CTC 21418
ACTACAACCT	CCA3CTCCCA	CGTTCAG33	ATTCTCATGC CTCAGTCTCC CGAGTAG3TG 21478
GSATTACAG3	CAT3CACCAC	TTACACCCGG	CTAATTTTTC TATTTTGTAG AGAGCTG3GG 21538
TTTCAACATG	TTG33CAG3C	T33TCTCAAA	CCCTTAACCT CAAGTGATCT GCCTGCCTCA 21598
GCCTCCCAAA	CAAAACAAAC	AC3CCACAGT	TTAATATGTG TTACAACACA CATGCTGCAA 21658
CTTTTATGAG	TATTTTAATG	ATATAGATTA	TAAAAGGTTG TTTTAACTT TTAAATGCTG 21718
GSATTACAG3	CAT3AGCCAC	TGTGCCAGGC	CTGAAGTGTG TTTTAAAAA TGTCTGACCA 21778
GCTGTACATA	GTCTCTTGCA	GACTGSCCAA	GTCTCAAAGT G3GAACAGGT GTATTAAGGA 21838
CTATCCTTTG	GTTAAATTTT	CGTAAATGTT	CCTGTGCAA3 AATCTTTCTA ACTAGACTTC 21898
TCATTTATTA	TATTTATTTT	AG AT AAT	GCA CCC CGG ACC ATA TTT ATT ATA 21949
	Asp Asn Ala Pro Arg Thr Ile Phe Ile Ile		
40	45		
AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG	GCT GTA ACT ATC TCT		21997
Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met	Ala Val Thr Ile Ser		
50	55	60	65
GTG AAG TGT GAG AAA ATT TCA ACT CTC TCC TGT	GAG AAC AAA ATT ATT		22045
Val Lys Cys Glu Lys Ile Ser Thr Leu Ser Cys	Glu Asn Lys Ile Ile		
70	75	80	
TCC TTT AAG GTAAGACTG AGCCTTACTT TGTTTTCAAT	CATGTTAATA TAATCAATAT		22103
Ser Phe Lys			
AATTAGAAAT ATAACATTAT TTCTAATGTT AATATAAGTA	ATGTAATTAG AAAACTCAAA		22163
TATCCTCAGA CCAACCTTTT GTCTAGAACA GAAATAACAA	GAAGCAGAGA ACCATTAAG		22223
TGAATACTTA CTAAAAATTA TCAAACTCTT TACCTATTGT	GATAATGATG GTTTTTCTGA		22283
GCCTGTGCACA GGGGAAGAGG AGATACAACA CTTGTTTTAT	GACCTGCATC TCCTGAACAA		22323
TCAGTCTTTA TACAAATAAT AATGTAGAAT ACATATGTGA	GTTATACATT TAAGAATAAC		22403
ATGTGACTTT CCAGAAT3AG TTCTGCTATG AAGAATGAAG	CTAATTATCC TTCTATATTT		22463
CTACAACCTT GTAAATTATG ATAATATTTT AATCC3TAGT	TGTTTTGTTG CTGATCCTTA		22523
GCCTAAGCTT TAGCACTAAG CTTCAGCTTC CAGTTGATGT	ATGTTATTTT TAAT3TTAAT		22583
CTAATTGAAT AAAAGTTATG AGATCAG3TG TAAAA3TAAT	GCTATATTA TCTTCAAGCC		22643
AG3TATAAAG TATTTCT3GC CTCTACTTTT TCTCTATTAT	TCTTCATTAT TATTTCTCTAT		22703
TATTTTTCTC TATTTCTCTC ATTATTGTTA GATAAACCAC	AATTAACAT AG3TACABAC		22763
T3AGCCAGTA AGAGTAGCCA G33ATGCTTA CAAATTTGCCA	ATGCTTCAGA GGAGAATTCC		22823
AT3TCAT3AA GACTCTTTTT GAGT33AGAT TT33CAATAA	ATATCCGCTT TCATGCCCAC		22883
CCAGTCCCCA CTGAAAGACA GTTAG3ATAT GACCTTAGTG	AAGGTACCAA G33GCAACTT		22943
G3TAG33AGA AAAAG33CAC TCTAAATAT AATCCAA3TA	AGAACAGTGC ATATGCAACA		23003
GATACAG3CC CCAGACAAAT CCCTCAG3TA TCTCCCTCCA	ACCAGAGTGC CACCCCTTCA		23063
G3TGACAATT TGGAGTCCCC ATTCTAGACC T33AG33CAG	CTTAGTTATC AAAATAG3AT		23123
AAGAG33CTG GGATG33AGG GTAG33T3GA AAG33TTAAG	CATGCTGTTA CTGAACAA3A		23183
TAATTAG3AG GGAAG33BAT G33CAAG3TC AAG33TATGT	GGATAGAGGA AAACTCAB3T		23243
G3AGAG33CAG ATTCAGAAAC TG33GATAAGT CC3AACTAC	AG3TGGATTC TTGTT3A333		23303
AGACT33T3A AAATGTTAAG AAGATG33AA TAATG3TTG3	CACTTAGTAG GAAGT333CA		23363
AATCCATATT T3GG3333CC T3AAGTTTAT TCAATTTTGA	T33CCCTTTT AAATAAAAA3		23423
AATGT333TG G3CGT33TG3 CTCACAC3TG TAATCCAG3C	ACTTTG33AG GCGAG33333		23483
GCGGATCAC3 TGAAGTCA33 AGTTCAAGAC CAG3CTGACC	AACATG33AG AACCC3AT3T		23543
CTACTAAAAA TACAAAATTA GTTG33G3TG GT33CATAT3	CCTGTAATCC CAG3TACTCG		23603
G3AGGCTGAG GCAG33AGAT CTTTTGAACC CG33AG33CAG	AG3T333GAT GAG33CTAGAT		23663
CGTGCCATTG CACTCCAG3C T3333CAACAA GAG3CAAACT	CG3TCTCAAA AAAAAAAA3A		23723
AAAAAGT3AA ATTAACCAAA G3CATTA3CT TAATAATTIA	ATACTGTTTT TAAGTA333C		23783
GGGGGGTG3C TGGAAGAGAT CT3T3TAAAT GAG33AATCT	GACATTTAAG CTTTATCA3C		23843
ATCATAG3AA ATCTGCTTCT G3AAG3AAT CAATAAATAT	TAGTTG33AG3 GGG33A3A3A		23903
GTGAGG33TG GACTAGGACC AGTTTTAG3C CTTGTCTTTA	ATCCCTTTTC CT33CACTAA		23963

TAAGGATCTT	AGCAGTGGTT	ATAAAAAGTGG	CCTAGGTTCT	AGATAATAAG	ATACAAACAGG	24023
CCAGGCACAG	TGGGTCATGC	CTATAATCCC	AGCACTTTGG	GAGGGCAAGG	CGAGTGTCTC	24083
ACTTGAGATC	AGGAGTTCAA	GACCAGCCTG	GCCAGCATG	CGATACTCTG	TCTCTACTAA	24143
AAAAAATACA	AAAATTAGCC	AGGCATGGTG	GCATGCACCT	GTAATCCCAG	CTACTCGTGA	24203
GCCTGAGGCA	GAAGAAATCG	TTGAAACCCAG	GAGGTGTAGG	CTGCAGTGAG	CTGAGATCGC	24263
ACCACTGGAC	TCCAGCCTGG	GCGACAGAAAT	GAGACTTTGT	CTCAAAAAAA	GAAAAAGATA	24323
CAACAGGCTA	CCCTTATGTG	CTCACCTTTT	ACTGTTGATT	ACTAGCTATA	AAGTCCTATA	24383
AAGTTCTTTG	GTCAGAAACC	TTGACAAACAC	TAAGAGGGAT	TTGCTTTGAG	AGGTTACTGT	24443
CAGAGTCTGT	TTCATATATA	TACATATACA	TGTATATATG	TATCTATATC	CAGGCTTTGG	24503
CAGGCTTCCC	TCACTACTTT	CAGTGCACCT	GGGAGATGTT	AGGTCAATAT	CAACTTTCCC	24563
TGSATTGAGA	TTCAACCCCT	TCTGATGTAA	AAAAAAGAAA	AAAAAAGAAA	GAAATCCCTT	24623
TCCCTTTGGA	GCACTCAAGT	TTCACCAGGT	GGGGCTTTCC	AAGTTGGGGG	TTCTCCAAAG	24683
TCATTGAGAT	TGCTTTTACA	TCCATTGCTT	ATGTACCTTT	CCTATGATGG	CTGGAGATGG	24743
TCAACATCAA	AAGTAGGAAA	GCTACTGCCC	AAGGATGTCC	TTACTCTAT	TCTGAAATGT	24803
GCAATAAGTG	TGATTAAAGA	GATTGCCTGT	TCTACCTATC	CACACTCTCG	CTTTCACTGT	24863
TAACCTTCTT	TTTTTCTTTT	TTTTTCTTTT	TCTTTTCTTT	TGAAACGGAG	TCTCGCTCTG	24923
TGCGCCAGGC	TAGAGTGCAG	TGGCACGATC	TCAGCTCACT	GCAAGCTCTG	CCTCCCGGCT	24983
TCAGGCTCATT	CTCCTGCTCT	AGCCTGCCAA	GCAGCTGGGA	CTACAGGCGC	CTGGCAACAT	25043
GCCCAAGTAA	TTTTTTGTAT	TTTTAGTAGA	GACGGGCTTT	CACCGTGTTA	GCCAGGATGG	25103
TCTCGATCTC	CTGAACCTGT	GATCGGCTCG	CCTCAGCTCT	CCAAAGTGCT	GGGATTACAG	25163
GCGTGAGCCA	TCCGACCCCG	CTCAACTGTA	ACTTTCTATA	CTGCTTCATC	TTCCCTGTGA	25223
ATGTTACTAG	AGCTTTTGAA	GTTTTGGCTA	TGGATTATTT	CTCATTATTA	CATTAGATTT	25283
CAGATTAGTT	CCAAATTGAT	GCCACAGCTT	TAGGCTCTCT	TCTTAAATTG	TATATTGTAG	25343
ACAGCTGCAG	AAGTGGGTGC	CAATAGGGGA	ACTAGTTTAT	ACTTTTCATC	ACTTAGGACC	25403
CACACTTGTT	GATAAAGAAC	AAAGGTCAAG	AGTTATGACT	ACTGATTCCA	CAACTGATTT	25463
AGAAGTTGGA	GATAACCCCG	TGACCTCTGC	ACTCCAGAGT	CTTTCAAGCA	TCTTTGAAAG	25523
ATGAAGAAAT	GCTATTTTAA	TTTTGGAGGT	TTCTCTATCA	GTGCTTAGGA	TCATGGGAAAT	25583
CTGTGCTGCC	ATGAGGGCAA	AATTAAGTCC	AAAACATCTA	CTGCTTCCAG	GATTAACATG	25643
GAAAGAACCT	AAGTGGTGCC	CACATGTTCT	GATCCATCCT	GCAAAATAGA	CATGCTGCAC	25703
TAACAAGAAA	AAGTGCAGGA	GCACTACCA	TTGATAAACC	TGCAAGATTA	TAGTTTCAAG	25763
TAATCTAAC	ATTTCTCACA	AGGCCCTATT	CTGTGACTGA	AACATACAAG	AATCTGATTT	25823
TGGCTTTCTA	AGGCAGGGCC	CAGCCAAGGA	GACCATATTC	AAGACAGAAA	TTCAAGACTA	25883
CTATGAAACT	GAGGTGCTTG	GCAGGGAAGA	CAGAGTCAA	GACTGCAAC	TGAGGCAATA	25943
CAGCAGCTTT	ACACAGGAAAC	CCAGGCTCTA	GCCCTACAAC	AATTATTGGG	TCTATTCACT	26003
GTAAGTTTAA	ATTTCAAGCT	CCACTGAAAG	AGTAAGCTAA	GATTCTGCTG	ACTTTCTGTC	26063
TCTCTCAGAG	TTGGCTCAGA	AATGAGAACT	GCTCAGGCGA	GCCATGCTGG	CTTACACTGT	26123
GAATCCCAGC	ACTTTGGGAG	GCCGAAGTGG	GAGGCTCACT	TGAGGCTCAG	AGTTCAAGGAC	26183
CAGCTTAAGC	AACAAAGTGA	GATACCCCTT	GACCCCTTCT	CTACAAAAAT	AAATTTTAAA	26243
AATTAGCCAA	ATGTGGTGGT	GTATACTTAC	AGTCCACAGT	ATTCAAGGAG	CTGAGGCTAG	26303
GGGATTGCTT	GAGCCAGGGA	ATTCAGGCTT	GCAATGAGCT	ATGATTTTAC	CAGTGTACTT	26363
CTGGCTGGGC	AACAGAGCGA	GACCCGTGCT	CAAGCAAAA	AGAAAAAGAA	ACTAGACTTA	26423
GCCTAAGTTT	GTGGGAGGAG	GTCATCATCG	TCTTTAGCCG	TGAATGGTTA	TTATAGAGGA	26483
CAGAAATTGA	CATTAGCCCA	AAAAGCTTGT	GCTCTTTGCT	GCAACTCTAC	TTAATCTTGA	26543
GCAAAATGTG	ACACCACTCA	ATGGAAGAGG	AGAGAAATTA	GCTGTTTGAT	GTATAGGGGA	26603
AAACTAGAGG	CCGTGAACTG	AATATGCTAT	CCATGAGAGG	GAGAAATAGG	GATTCGAGGT	26663
TAAGAAAGAG	AGGAGGTGAG	TACTGCTGTT	CAGAGATTTT	TTTTATGTAA	CTCTTGAAGAA	26723
GCAAAACTAC	TTTTGTTCTG	TTTGTAATA	TACTTCAAAA	CAAACTTCAT	ATATTCAAAAT	26783
TGTTTATGTC	CTGAAATAAT	TAGGTAATGT	TTTTTTCTCT	ATAG GAA ATG AAT CTT		26839

Glu Met Asn Pro

85

CCT	GAT	AAC	ATC	AAG	GAT	ACA	AAA	AGT	GAC	ATC	ATA	TTC	TTT	CAG	AGA	26887
Pro	Asp	Asn	Ile	Lys	Asp	Thr	Lys	Ser	Asp	Ile	Ile	Phe	Phe	Gln	Arg	
90						95					100					
AGT	GTT	CCA	GGA	CAT	GAT	AAT	AAG	ATG	CAA	TTT	AAA	TCT	TCA	TCA	TAC	26935
Ser	Val	Pro	Gly	His	Asp	Asn	Lys	Met	Gln	Phe	Glu	Ser	Ser	Ser	Tyr	
105					110				115						120	
GAA	GGA	TAC	TTT	CTA	GCT	TGT	GAA	AAA	GAG	AGA	GAC	CTT	TTT	AAA	CTC	26983
Glu	Gly	Tyr	Phe	Leu	Ala	Cys	Glu	Lys	Glu	Arg	Asp	Leu	Phe	Lys	Leu	
				125				130							135	
ATT	TTG	AAA	AAA	GAG	GAT	GAA	TTG	GCG	GAT	AGA	TCT	ATA	ATG	TTC	ACT	27031
Ile	Leu	Lys	Lys	Glu	Asp	Glu	Leu	Gly	Asp	Arg	Ser	Ile	Met	Phe	Thr	
			140					145					150			
GTT	CAA	AAC	GAA	GAC	T	AGCTATTAAA	ATTTTCATGCC	GGGCGAGTG	GCTCACGCCCT							27087
Val	Gln	Asn	Glu	Asp												
155																

GTAATCCCAG	CCCTTTGGGA	GCGTGAGGCG	GCCAGATCAC	CAGAGGTCAG	GTGTTCAAGA	27147
CCAGGCTGAC	CAACATGGTG	AAACCTCATC	TCTACTAAAA	ATACAAAAAA	TTAGCTGAGT	27207

GTAGTGACCC	ATGCCCTCAA	TCCCAGCTAC	TCAAGAGGCT	GAGGCAGGAG	AATCACTTGC	27267
ACTCCG3AGG	TGGAGGTTGT	GGTGAGCCGA	GATTGCACCA	TTGCGCTCTA	GCCTGGGCAA	27327
CAACAG3AAA	ACTCCATCTC	AAAAAATAAA	ATAAATAAAT	AAACAAATAA	AAAATTCATA	27387
ATGTGAACTG	TCTGAATTTT	TATGTTTGA	AAGATTATGA	GATTATTAGT	CTATAATTGT	27447
AATGGT3AAA	TAAAATAAAT	ACCACTCTTG	AAAAACATCA	TTAAGAAATG	AATGAACTTT	27507
CACAAAAGCA	AACAAACAGA	CTTTCCCTTA	TTTAAGTGAA	TAAAATAAAA	TAAAATAAAA	27567
TAATGTTTAA	AAAATTCATA	GTTTGAAAAC	ATTCTACATT	GTTAATTGGC	ATATTAATTA	27627
TACTTAATAT	AATTATTTTT	AAATCTTTTG	GGTTATTAGT	CCTAATGACA	AAAGATATTG	27687
ATATTTGAAC	TTTCTAATTT	TTAAGAATAT	CGTTAAACCA	TCAATATTTT	TATAAGGAGG	27747
CCACTTCACT	TGACAAATTT	CTGAATTTCC	TCCAAAGTCA	GTATATTTTT	AAAATTCAGT	27807
TTGATCCTGA	ATCCAGCAAT	ATATAAAAGG	GATTATATAC	TCTGGCCAAC	TGACATTCAT	27867
CCTAGGAATG	CAAAGATGGT	TTAATATCCT	AAAATCAATT	AACATAACAT	ACTATATTAA	27927
TAAAGTATCA	AAACAGTATT	CTCATCTTTT	TTTCTTTTTT	CACAATTCCT	TGGTTACACT	27987
ATCATCTCAA	TAGATGCAGA	AAAAGCATTT	GACAAAATCC	AATTCATAAT	AAAAATTCTC	28047
AAACTT3AAA	GAGAACATCA	TAAAG3CATC	TATGAAAAAC	CTACAGCTAA	TATCATACTT	28107
AACGATGAAA	AACTGAATTA	TTTTACCCTA	AGATCAAGAA	TAATGCAAGC	ATGTCAGCTC	28167
TTGCAACTTC	TATTCAACAT	TGTACTGGAG	GTTCTAGCCA	GAGCAACCAT	ACAATAAATA	28227
AAAATAAAAG	GCA3CCAGAT	TAGAAAGGAA	GTCTTTATTT	GCAGACAACA	TGGTTCTTTA	28287
TGCAGAAAC	CGTCAGGAAT	ACACACACAT	GTTAGAACTA	ATAAGTTCAG	CAAGSTTGCA	28347
GGTTGCAATA	TCAATATGCA	AAATACATT	GAAGGCTGGG	CTCAGTGGAG	ATGGCATGTA	28407
CCTTTCTGTC	CAGCTACTTG	GGAGGCTGAG	GTAGGAGGAT	CACTTGAGGT	GAGGAGTTTG	28467
AGGCTATAGT	GCAATGTGAT	CTTGCTGTG	AATAGCCACT	GCACTCGAGC	CTAGGCAACA	28527
AAGTGAGACC	CCGTCTCCAA	AAAAAAAAT	GGTATATTGG	TATTTCTGTA	TATGAACAAT	28587
GAATGATCTG	AAAACAAGAA	AATTCCATTC	ACGATGGTAT	TAAAAAATA	AAATACAAAT	28647
AAATTTAGCA	AAATAATTAT	AAAACCTGTA	CATCGAAAAAT	TTCAAAGCAC	TCTGAGGGAA	28707
ATTAAAGATG	ATCTAAATAA	TTGAGAGAGC	ACTCTATGAT	CACTGATTGG	AAAATTCATT	28767
CAATATTGTT	AAGATAACAA	TTGTCCCAAA	ATTGATGCAT	GCATTCAATT	TAGTCTTCAT	28827
CAAAATTCCA	GCAGGGTTTT	TGCAGAAATT	GACAAGCTGT	ACCCAAAATG	TATATGGAAA	28887
TGAAAAGACC	CAGAAGAGCA	AATAATTTTT	TAAAAACAAA	GTTGGAAAAC	TTTACTTCC	28947
TAATTTTAAA	ACTTACTATA	AACTAAAGT	TATCAAGACC	ATTTAGT		28994

(15) INFORMATION FOR SEQ ID NO: 15:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (v) FRAGMENT TYPE: N-terminal fragment
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

Tyr	Phe	Gly	Lys	Leu	Glu	Ser	Lys	Leu	Ser
1				5					10

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 27 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CCATCCTAAT ACGATTCACAT ATAGGGC

27

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 28 base pairs

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TTCCTCTTCC CGAAGCTGTG TAGACTGC

28

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CTATAGGGCA CGCGTGST

18

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 28 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TTCCTCTTCC CGAAGCTGTG TAGACTGC

28

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GTAAGTTTTC ACCTTCCAAC TGTAGAGTCC

30

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GGGATCAAGT CGTGATCAGA AGCAGCACAC

30

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CCTGGCTGCC AACTCTGGCT GCTAAAGCGG

30

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 33 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GTATTGTCAA TAAATTCAT TGCCACAAAG TTG

33

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 33 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

AAGATGGCTG CTGAACCACT AGAAGACAAT TGC

33

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 27 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TCCTTGGTCA ATGAAGAGAA CTTG3TC

27

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 33 base pairs

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CCTGGAATCA GATTACTTTG GCAAGCTTGA ATC

33

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 32 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GGAAATAATT TTGTTCTCAC AGGAGAGAGT TG

32

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 31 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GCCAGCCTAG AGGTATGGCT GTAACATCT C

31

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GGCATGAAAT TTTAATAGCT AGTCTTCGTT TTG

33

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GTGACATCAT ATTCTTTCAG AGAAGTGTCC

30

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 31 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GCAATTTGAA TCTTCATCAT ACGAAGGATA C

31

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 31 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

TCCGAAGCTT AAGATGGCTG CTGAACCACT A

31

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 32 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GGAAATAATT TTGTCTCAC AGGAGAGAGT TG

32

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 37 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

ATGTAGCGGC CGCGGCATGA AATTTTAATA GCTAGTC

37

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 33 base pairs

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CCTGGAATCA GATTACTTTG GCAAGCTTGA ATC

33